

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Hox gene regulation is central to the evolution of five-digit tetrapods from polydactyl ancestors **PAGE 89**



DIGITAL ARRAY

ENERGY

AFRICA'S POWER STRUGGLE

Can a continent bypass fossil fuels and adopt renewables?

PAGE 20

SCIENTIFIC METHOD

ONE THEORY IS NOT ENOUGH

There may be another explanation — go look for it!

PAGE 23

COLLEGE EDUCATION

WHAT'S ON THE SYLLABUS?

An open-access resource celebrates ways of learning

PAGE 125

NATURE.COM/NATURE

3 November 2016 £10

Vol. 539, No. 7627



9 770028 083095

THIS WEEK



EDITORIALS

GET REAL! The realities of a research life escape some **p.6**

WORLD VIEW Geneticists must return volunteer gene data **p.7**

SAFETY Roads make mongooses stop looking and listening **p.9**

Trapped in a bear hug

Continuing uncertainty in Russia is bad for science and the wider world — and Putin's stifling control of key appointments and private initiatives is unlikely to reverse decades of post-Soviet decline.

As the United States goes to the polls after an unprecedentedly vulgar election campaign, its former cold-war rival is exuding strength and confidence. Just last week, Russia cited its proud history of exploration and science in Antarctica when giving the green light to an international agreement to create a vast marine reserve in the Ross Sea (see page 13). Even so, Vladimir Putin's chauvinistic approach to politics has a certain appeal to some — at home and abroad.

The Russian leader's perplexing popularity in some circles detracts from his enduring failure to modernize his country's government, society and economy. Russia's flagging science system is in dire need of a cure, too. But the recent appointment of Olga Vasilyeva as Putin's science and education minister casts strong doubts on whether the right healers are at work. Vasilyeva, an ultra-conservative historian, is mainly known for her affinity with the Russian Orthodox Church and her ambivalence towards Stalin.

Russian science is still struggling to recover from decades of neglect and post-Soviet degradation. Its research community is isolated. Foreign students and scientists are a rare commodity at Russian universities and research institutes. A generously funded government scheme to attract top foreign researchers to Russian labs is hampered by red tape: Westerners attempting collaborative science in Russia often complain that the security services and customs authorities interfere with civil research — harassments that many Russian scientists have been quietly enduring for many years.

According to official reading, science is held in high esteem. Putin and the clique of allies he has placed in key positions in industry and administration like to think of Russia as a technological powerhouse. But government programmes such as RUSNANO, a multibillion-rouble nanotechnology initiative launched in 2007, have by all accounts delivered little in terms of innovation. The Skolkovo Innovation Centre outside Moscow, hailed as a Russian Silicon Valley, grapples with allegations of embezzlement against members of its management body. Private initiatives and philanthropic ventures, meanwhile, are feeling the effects of jealous state control: the Moscow-based Dynasty Foundation, a rare private science-funding body, ceased operations last year after it was labelled an undesired 'foreign agent'.

Even so, some growth in public science spending in recent years, together with plans to strengthen universities and streamline the oversized Russian Academy of Sciences (which runs hundreds of research institutes), raised hopes that things might start to improve. Dmitry Livanov, a dynamic physicist who ran the science ministry from 2012, had seemed the right person to push through a series of reforms to get Russian science back on course. It came as a shock, therefore, when Putin fired him in August.

The reasons for Livanov's abrupt departure remain unclear, although they are likely to be political rather than related to the performance of the ministry he headed. His dismissal might be considered a win for the Russian Academy of Sciences and for numerous

low-profile universities, which he intended to close. But wins in these cases would not be a win for Russian science at large.

The arrival of his successor has created fresh uncertainty. In one of her first moves, on 28 September, Vasilyeva announced her intention to suspend planned university mergers and expressed doubts about the future of a government programme to create five world-class universities by 2020. Then last month she said that Russian scholars and scientists should be assessed primarily on the basis of their publications in Russian-language academic journals. Scientists were puzzled: would the proposed system apply only to the humanities or to all fields of research, in which case it would render much of Russian science a footnote in global terms? Her ministry has failed to respond to requests for clarification from *Nature*.

Livanov's removal and Vasilyeva's awkward manoeuvres during her first months in office indicate that science and science-related affairs are becoming increasingly subject to Putin's autocratic scheming. Moscow's suspension last week of an agreement with the United States on cleaning up weapons-grade plutonium fits with that view.

But no matter how tense the geopolitical climate, Putin must understand that isolationism leads to a dead end in both science and politics. Livanov's reforms should continue, and Vasilyeva must urgently provide Russia's anxious research community with a clear outlook. Putin and his ministers must also strive for more constructive international collaboration, in science and in other spheres. Russia cannot go it alone, either in science or in Syria. One can only hope that its consent to join efforts to protect the high seas might herald a new era. ■

"Putin must understand that isolationism leads to a dead end in both science and politics."

Spots and stripes

Deciphering the genes at play in the patterning of mammalian coats.

The Argentine writer Jorge Luis Borges tells the story of an Aztec priest, Tzinacán, imprisoned by the conquistador Pedro de Alvarado. In the cell next door is a jaguar: Tzinacán becomes convinced that the jaguar's spots are not random blotches, but contain a message from his God that, could he decipher it, would offer a key to his escape. Any reader of Borges learns to appreciate his playful mash-up of fact and fiction. Of the small cast of characters in *La Escritura de Dios* (*The God's Script*), de Alvarado was a real person, and Tzinacán probably an invention, although with Borges one can never be sure.

Jaguars, though, definitely exist, and — like many mammals — have a pattern of spots that fascinates and tantalizes.

Understanding the origins of variegated colour patterns in mammalian fur is an abiding problem in biology. Other animals adopt a range of pigments, and even use optical effects such as iridescence to lend a chromatic gloss, yet the mammalian palette is mainly monochrome. A patch of skin either contains melanocytes, or it doesn't.

This week, researchers report in *Nature* some progress on the problem with the African striped mouse, *Rhabdomys pumilio* (R. Mallarino *et al.* *Nature* <http://dx.doi.org/10.1038/nature20109>; 2016). This creature has a stripe on either side of its spine, each a sandwich of light-coloured hair between two outriders of pure black. The rest of the mouse is an intermediate shade, except for a pale belly. The pattern starts to emerge long before a mouse is born.

The difference is down to gene expression. The white stripes are enriched in transcripts of *Alx3*, a transcription factor, which curbs the activities of a gene called *Mitf*. If left unhindered, this gene would allow melanocytes to differentiate and produce dark pigment.

As model organisms go, *R. pumilio* is very different from the laboratory mouse. Even further removed is the Eastern chipmunk, *Tamias striatus*. Chipmunks are more closely related to squirrels than to mice: the last common ancestor of mouse and chipmunk lived when dinosaurs did. Yet the formation of chipmunk stripes is governed by essentially the same processes that create the patterning in mouse skin, even though the mechanisms might have evolved independently in each case.

Study of the chipmunk shows other genes involved. Expression of one called *Asip* in lighter areas, another called *Edn3* in darker, show that patterning is not down to a single genetic interaction. The work of *Edn3* and other genes, we know, writes the script of spots and stripes in cats, from tabbies to cheetahs (C. B. Kaelin *et al.* *Science* **337**, 1536–1541; 2012) — and so, presumably, in the coat of the jaguar that Tzinacán longed to decipher.

Much remains to be learnt. The stripes of mice and chipmunks don't occur in the same places on the animal, and scientists still do not understand why the grass mouse *Lemniscomys rosalia* has only one stripe, whereas the ground squirrel *Ictidomys tridecemlineatus* has thirteen. The God's script comes in many dialects.

Skin pigmentation is superficial — literally — but the genes that create these patterns often have other, more profound purposes. The skin and hair of vertebrates derives from the neural crest, an embryonic tissue unique to vertebrates, which, migrating from the edge of the neural

“Oddities of skin pigmentation sometimes betoken deeper ailments.”

plate as it rolls up to create the spinal cord, interacts with tissues all over the body to create structures seen nowhere else in the kingdom of life. The neural crest sculpts not just hair, teeth and skin, but a long list of attributes, from the bones of the face to the nerves that line the intestines, parts of the heart and adrenal glands, and many crucial components of our sense organs. This is why oddities of skin pigmentation sometimes betoken deeper ailments.

It explains why cats that are white are more than usually likely to be deaf.

So much is clear for *Alx3*. Mice deficient in this gene show a range of neural-tube closure defects, the incidence of which is reduced by folic acid (S. Lakhwani *et al.* *Dev. Biol.* **344**, 869–880; 2010). This may explain why human mothers deficient in this vitamin run the risk of giving birth to babies with spina bifida. Again in humans, recessive mutations in *ALX3* produce a series of facial malformations called frontorhiny, also related to failure of the facial bones to knit properly (S. R. F. Twigg *et al.* *Am. J. Hum. Genet.* **84**, 698–705; 2009). The script runs deep, with many layers of meaning.

Did Tzinacán finally decipher the God's script? The answer is yes: the jaguar's fur encoded a spell which, if recited out loud, would make the prison vanish. But Tzinacán chose not to use it because, in the act of decipherment, he became a god himself. ■

Get real

Researchers must show policymakers that scientific evidence is far from academic.

Grammar and Twitter rarely sit happily, so it would be churlish to point out that when the Welsh MP Glyn Davies tweeted at the weekend: “Nothing more irritating than academics rubbishising the efforts of those operating at the sharp end, without facing up to the hard decisions”, he was inadvertently complaining that people at the sharp end (himself included presumably) do not confront hard decisions.

Besides, the next social-media missive from Davies made his position clear: “Personally, never thought of academics as ‘experts’. No experience of the real world.” His first point there might — just — be semantically defensible: academics, by one definition, are full-time scholars; whereas experts can be classed as those who have learned not through study but through experience. But it was his second assertion that prompted most of the angry backlash, and the inevitable hashtag response #realworldacademic that was still going strong as *Nature* went to press.

(Replies to Davies ranged from “Practiced medicine in Intensive Care Unit and emergency medicine while I was doing a PhD” to “Dude, you literally work in a palace.”)

There's no need for *Nature* to tell its readers — mostly academics — that they have experience of the real world. They live it every day; and, for many, the realities of this academic life are starting to bite down hard. As we explored in a special issue last week, the real world of academia for many young researchers is insecure and under increasing

pressure. Many are looking to leave. (And when they do, Davies and others please note, they seem to flourish.)

The popular image of an academic as aloof, privileged and out of touch — if it ever was true — is now redundant. But then so is the popular view that backbench MPs are, well, aloof, privileged and out of touch. In most cases, both groups work harder, and with more selfless goals, than critics claim. By their nature, those who study the science of what is probable will come into conflict with those who practise the art of what is possible. But researchers, along with everybody else who criticizes policymakers and elected officials, should remember that, as Davies seemed to be trying to point out, it is one thing to discuss problems and recommend solutions, and quite another to have to make and implement decisions.

One reason that the MP's comments seem to have struck a nerve is that they feed into the popular idea — fuelled by the Brexit campaign and the rise of Donald Trump — that politicians, and by extension the wider public, have shifted away from reason and evidence. In a recent World View column, Bill Colglazier, a former science adviser to the US government, argued that this perception could be explained by differing attitudes to evidence — and on this point researchers seem to have some common ground with Davies.

Criticized last month for attending a lecture by the prominent climate sceptic Matt Ridley at the prominent climate sceptic organization the Global Warming Policy Foundation, Davies wrote on his blog: “I do not think Government policy should be based on a partial view of science. I like to make judgements based on evidence ... In the end, governments the world over will be guided by evidence — or science delivered as evidence.”

The conflict between Davies' support for evidence and his Twitter dismissal of those who seek and provide evidence seems, in the real world, to make for a curious paradox. Perhaps an expert could look into it. ■

DEREK ARMSTRONG



Geneticists should offer data to participants

Sarah Nelson was refused access to her own genome data. How long before volunteers who face this attitude turn away from science?

As a human-genetics researcher, I analyse the DNA of thousands of anonymous strangers. Earlier this year, I got to experience the other side of a consent form — and was left disappointed. When another research group asked me to donate my own genetic material for their whole-genome sequencing project, I asked in exchange for access to my raw data — to explore, to play with and just to have on file. Not surprisingly, my request was refused: the status quo for biomedical and genetic studies is not to return individual-level data to participants.

I still joined the study, but the irony is not lost on me that my personal data will be available to thousands of scientists (including me) through restricted-access databases. As awareness and usefulness of this information increases, I fear that potential volunteers who are refused access to their genetic data will become less willing to donate them to science. The genetics-research community must therefore update its stance on returning personal data.

Granted, there are well-founded reasons why studies don't typically return the data. Researchers rarely recontact participants, and doing so could draw resources and attention away from the primary project goals. Research is not medicine, and returning data can create the misleading impression that researchers are offering health care. Historically, there has also been little reason to return genetic data, because volunteers couldn't access tools to receive, store or understand them.

But beyond the lab, more people now want and expect access to all kinds of personal data, a trend that shows no signs of slowing. Health data are no exception, as evidenced by the flood of wellness and 'mobile health' apps that are now coming to market. Future generations will take for granted that our personal computing devices are vehicles for almost unlimited 'quantified self' and self-tracking activities.

Several online platforms exist to help people to explore their genetic data, developed by for-profit companies, academic groups or by self-taught citizen scientists. Since launching in 2011, the site openSNP has drawn more than 4,000 users, half of whom have uploaded genetic data. The DNA.LAND platform has attracted more than 32,000 contributors since its release last October. Launched in 2008, and therefore one of the earliest third-party interpretation tools, Promethease reports performing hundreds of analyses daily. Other tools, such as GEDMatch and Genome Mate Pro, attract thousands of users who are eager to incorporate genetic analyses into their genealogical research.

Many scientists are suspicious and occasionally derisive of consumer or 'recreational' genomics. Although these products have their flaws, they underscore what citizens can and want to do with their genetic data. For the genetics-research community to maintain its good relationship with volunteers, it must take these activities more seriously. Many current large-scale genetics-research studies rely on legacy collections, and have not had to navigate the new 'participant-as-owner'

culture. But legacy studies cannot fuel future research indefinitely.

The genetics-research community needs to develop an anticipatory infrastructure to return raw data to interested participants. The format for returning genotype data would probably vary according to the nature of the study — depending, for instance, on whether next-generation sequencing or microarray genotyping was performed. Funding opportunities should include resources for participant data return. Institutional review boards need to be able to review the mechanisms that studies propose for returning genetic data. Research groups should develop and adopt informed-consent procedures so participants can make decisions about acquiring their raw data, including the limitations of self-directed interpretation and analysis.

Making data-return practicable might require building technical systems such as secure web interfaces. However, we already have secure and robust methods to share data within the scientific community, so perhaps the necessary change in culture is a bigger hurdle.

Some research initiatives are already experimenting. In early September, the New York Genome Center in New York City released Seeq, a smartphone app and research platform through which individuals can pay a modest fee (around US\$50) to receive their whole-genome sequence and some interpreted reports (such as ancestry composition and microbiome profiles). In turn, the researchers amass genomes for their own research projects.

Most of the available consumer-genomics tests look at a very small portion of the genome, but newer and more powerful exome and whole-genome sequencing options offer more detail at a rapidly decreasing cost. MyGene2, a web portal created by researchers at the University of Washington in Seattle, enables the sharing of genetic-sequencing and medical data across families, physicians and researchers, with the goal of tackling rare genetic diseases. Genomics tools such as this flourish when people are stewards of their own genetic data.

This era of big data begs big questions, including who should own health and research data. Legally, research participants may not own their biological specimens, or the data extracted from them, once these have been donated to scientific studies. But as researchers, don't we have an obligation to respect the individual autonomy of participants seeking their raw data? Asking for access is not the same as asking for ownership or control, just for a reasonable reciprocity. Let's have the conversation. After all, if potential participants can obtain their genetic data from a growing number of commercial companies, they might turn their backs on traditional research studies altogether. ■

Sarah Nelson is a researcher and PhD student at the Institute for Public Health Genetics at the University of Washington in Seattle. e-mail: sarahcn@uw.edu

GENOMICS TOOLS
FLOURISH
WHEN PEOPLE ARE
STEWARDS OF THEIR
OWN GENETIC
DATA.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PLANT NANOTECHNOLOGY

Bionic plant can sense explosives

By incorporating fluorescent carbon nanotubes into spinach plants, researchers have turned the plants into environmental sensors.

Michael Strano and his colleagues at the Massachusetts Institute of Technology in Cambridge coated carbon nanotubes with a peptide that binds to nitroaromatic compounds, which include explosives. They embedded the nanoparticles into the leaves of spinach plants. When chemical contaminants are absorbed by the roots or leaves, they attach to the nanotubes, causing the nanotubes' fluorescence to decrease by an amount that depends on the level of the compound. A small detector picks up the signal and relays it wirelessly to a smartphone.

Living-plant sensors could be deployed to large, remote areas for chemical monitoring, the authors say.

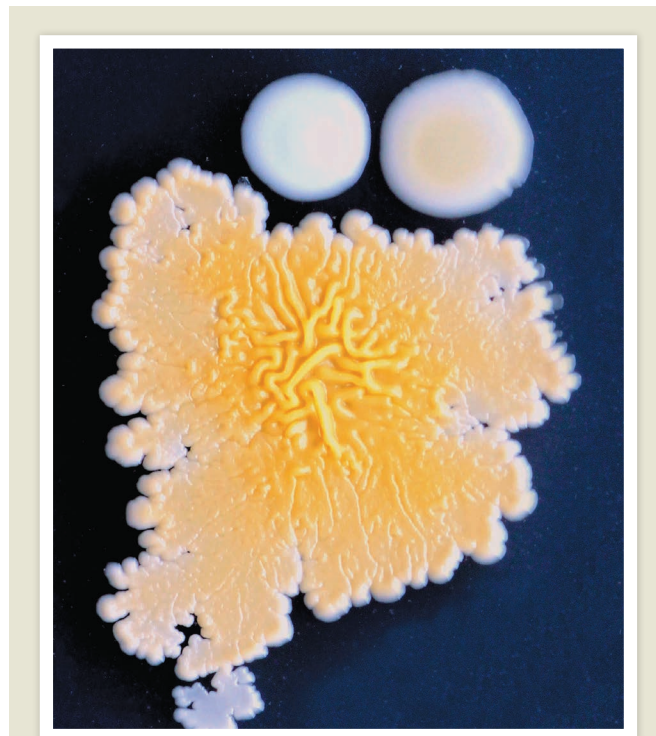
Nature Mater. <http://dx.doi.org/10.1038/nmat4771> (2016)

METABOLISM

Low oxygen resets the body clock

Cutting ambient oxygen levels helps mice to recover from a situation similar to jet lag.

In mammals, circadian clocks synchronize metabolism according to the day–night cycle. Gad Asher at the Weizmann Institute of Science in Rehovot, Israel, and his colleagues found that the amount of oxygen in the blood and kidneys of rodents varies with the time of day. Tests in cultured mouse cells showed that rhythmic fluctuations in oxygen levels synchronized



MICROBIOLOGY

Fungi boost bacterium

A study of 25 cheeses finds that a slow-growing bacterium can outcompete its relatives with the help of fungi.

Benjamin Wolfe at Tufts University in Medford, Massachusetts, and his colleagues examined the relative abundance of *Staphylococcus* bacteria (three species pictured), which are common in cheese. They found that *Staphylococcus equorum* dominated, despite being the slowest grower in lab tests. In the presence of fungi of the genus *Scopulariopsis*, *S. equorum* lowered its expression of genes involved in iron uptake and metabolism. The fungi could be providing the bacterium with freely available iron needed for growth, saving *S. equorum* the effort of acquiring and processing the nutrient, and allowing it to outcompete other bacteria.

Fungi could be influencing the diversity of other bacterial communities, including those in humans, the authors say.
mBio 7, e01157-16 (2016)

the circadian clock; this seemed to happen through HIF1 α , a protein known to be an oxygen sensor. Mice exposed to a cycle of light and dark that was shifted by six hours to mimic jet lag adapted faster to the new conditions

when ambient oxygen levels were decreased either before or after the shift.

Modulation of oxygen levels could be a future therapy for jet lag, say the authors.

Cell Metab. <http://doi.org/bsc9> (2016)

ASTRONOMY

Small stars host water worlds

Earth-sized planets covered in water may be abundant around red dwarfs, the most common type of star in the Universe.

Yann Alibert and Willy Benz at the University of Bern used computer simulations to predict the properties of planets that could form around red dwarfs and host liquid water. They found that the radius of the planets would be 0.5–1.5 times that of Earth, with most being around the same size as Earth. More than 90% of the simulated planets were at least 10% water by mass, suggesting that they were completely surrounded by deep oceans.

The authors say that the prospects for life on such planets are unclear, because too much water could destabilize the climate.

Astron. Astrophys. in the press;
Preprint at <https://arxiv.org/abs/1610.03460> (2016)

IMMUNOLOGY

Weary T cells may not recover

Exhausted immune cells bear distinct genetic signatures, and may be difficult to revive — a finding with implications for therapies that harness the cells.

Immune cells called T cells can become 'exhausted' and dysfunctional after exposure to cancer or chronic infection. Two teams — one led by John Wherry at the University of Pennsylvania in Philadelphia, the other by Nir Yosef at the University of California, Berkeley, and Nicholas Haining at the Dana-Farber Cancer Institute in Boston, Massachusetts — looked at

E. K. KASTMAN ET AL./MBIO (CC BY 4.0)

SANGKHOH HUNGKHUNHOD/GETTY

changes in gene expression and epigenetic markers (chemical changes to DNA that do not affect its sequence) in mice infected with a virus. They found that exhausted T cells had a characteristic profile that distinguished them from functional T cells.

One of the teams also showed that exhausted T cells were reactivated by an antibody that blocks PD-L1, a protein that suppresses T-cell responses. However, this effect was transient when viral levels remained high, suggesting that certain kinds of immunotherapy may need to be combined with other treatments to yield lasting benefit.

Science <http://doi.org/bsdh>; <http://doi.org/bsdj> (2016)

ANIMAL BEHAVIOUR

Noise disrupts other senses

Noise pollution can affect how wild animals respond to other sensory inputs, such as smell.

Andrew Radford and his colleagues at the University of Bristol, UK, studied the behaviour of wild dwarf mongooses (*Helogale parvula*; pictured) that had been habituated to the presence of human observers. The team placed faeces from either a predator or a herbivore outside the mongoose burrow. When ambient natural sounds were played, mongooses were quick to inspect both types of faeces. In response to predator faeces, the animals showed increased vigilance and stayed close to the burrow. By contrast, when road noise was played, mongooses were slower to approach and showed similar

responses to both predator and herbivore faeces.

Noise pollution may distract the mongooses and increase stress, impairing the creatures' natural anti-predator behaviour, the authors say.

Curr. Biol. 26, R911–R912 (2016)

MATERIALS

3D-printed device shapes ultrasound

A specially designed lens can create ultrasound beams with the potential to precisely move, manipulate and destroy cell-sized objects.

Ultrasound beams can be made by firing pulses of laser light at a lens to create high-frequency vibrations. But glass lenses can create only relatively simple wave patterns. Claus-Dieter Ohl and his colleagues at Nanyang Technological University in Singapore used a 3D printer to build polymer lenses in 3D curved shapes. These lenses generated beams just as powerful as those made from glass, but their complex shapes allowed greater control over the beam's focus in space and time.

This could enable complex manipulations of minuscule objects, say the authors. **Appl. Phys. Lett.** 109, 174102 (2016)

FLUID DYNAMICS

Soft surfaces suppress splash

Splashing occurs when droplets strike a stiff, flat surface, but a soft material, such as silicone gel, can reduce or even eliminate splatter.

A team led by Robert Style



at the Swiss Federal Institute of Technology in Zürich and Alfonso Castrejón-Pita at the University of Oxford, UK, observed ethanol drops falling onto silicone gels of varying stiffness. Deformations of the soft substrates within a few microseconds of impact absorbed the drops' kinetic energy, decreasing splashing.

The authors say soft gels and elastic polymers could be used as inexpensive coatings to prevent splashing, which could improve many technologies, including inkjet printers.

Phys. Rev. Lett. 117, 184502 (2016)

ECOLOGY

River fish feed millions

Total freshwater-fish consumption provides for the dietary animal-protein needs of the equivalent of 158 million people, with poorer nations especially dependent on this natural and inexpensive source of food.

Peter McIntyre at the University of Wisconsin, Madison, and his colleagues used data from the Food and Agriculture Organization of the United Nations to build a global map of river fisheries, which have historically received less attention than their marine counterparts. They found that pressure from fishing was most intense in areas where biodiversity was also highest, raising concerns about conservation. The

Mekong (pictured), Amazon and Niger were some of the most heavily fished rivers, whereas rivers in the United States and Europe saw lower than expected catches.

Declines in river fish could be catastrophic for the food security of hundreds of millions of people, the authors say.

Proc. Natl Acad. Sci. USA <http://doi.org/bscf> (2016)

ANIMAL BEHAVIOUR

Magpies behave cooperatively

A species of magpie is the first bird found to show cooperative behaviour without prompting.

A team led by Lisa Horn at the University of Vienna devised apparatus that allowed East Asian azure-winged magpies (*Cyanopica cyana*) to distribute food (mealworms and crickets) to others and found that they gave out food relatively evenly to group members. The authors argue that their findings support the 'cooperative breeding hypothesis'. This states that prosocial behaviour — helping others at no or low cost — evolved in species such as humans, whose offspring are cared for by not only parents, but also other group members.

Biol. Lett. 12, 20160649 (2016)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

JOEL SARTORE/GETTY



SEVEN DAYS

The news in brief

EVENTS

Antarctic reserve

In a diplomatic breakthrough, 24 nations and the European Union agreed on 28 October to create the world's largest marine reserve, in the Southern Ocean. The deal, which will take effect in December 2017, protects 1.55 million square kilometres of the Ross Sea, a deep Antarctic bay 3,500 kilometres south of New Zealand, from commercial fishing and mineral exploitation. The agreement became possible because of assent from Russia, which had long blocked a deal. See page 13 for more.

Cholera in Haiti

On 27 October, health officials announced plans to vaccinate 820,000 people against cholera in regions of southern Haiti that were devastated by Hurricane Matthew last month. Damage to water and sanitation infrastructure has raised the risk of a cholera outbreak. In 2010, after a devastating earthquake, the country suffered an epidemic of the disease that affected some 700,000 people, killing around 9,000. Vaccinations will begin on 8 November, according

NUMBER CRUNCH

193 km

The distance covered by the first commercial journey of a self-driving truck. Transport firm Otto, owned by Uber, delivered 51,744 cans of Budweiser beer from Fort Collins, Colorado, to Colorado Springs.

Source: Otto



MASSIMO PERCOSSI/ANSA/AP

Italy hit by strongest quake in decades

A magnitude-6.6 earthquake struck central Italy on 30 October, the most powerful such event in the country since 1980. The epicentre was about 115 kilometres northeast of Rome. The town of Arquata del Tronto suffered heavy damage and Norcia's cathedral was destroyed, but no fatalities were reported. Italy's civil protection agency said that more than 15,000 people are in temporary

accommodation. The quake follows a series of tremors last week, and many towns in the region had already been evacuated following the magnitude-6.2 earthquake that hit the same area on 24 August and killed nearly 300 people, many of them in Amatrice (pictured). Geophysicists have been concerned about continuing activity in the region's complex system of faults.

to the Pan American Health Organization, which will support the campaign led by the Haitian Ministry of Health.

Mosquito test

The world's biggest test yet of an unconventional but promising method to fight mosquito-borne diseases will commence in Rio de Janeiro, Brazil, and Medellín, Colombia, scientists announced on 26 October. Mosquitoes that carry *Wolbachia* bacteria — which hinder the insects' ability to transmit Zika, dengue and other viruses — will be widely released in the cities over the

next two years, reaching an estimated 2.5 million people in each region. See page 17 for more.

PHARMACEUTICALS

EU drug agency

Ireland's health minister, Simon Harris, announced on 25 October that the country will formally bid to host the European Medicines Agency (EMA), which is widely expected to relocate from its base in London after Britain leaves the European Union. The EMA is responsible for the evaluation and approval of drugs marketed in the EU. Ireland and Spain have both

previously said that they would like to host the regulator, but Ireland's announcement is the most formal statement of intent so far. Harris said that he believes a move to Dublin would allow the EMA to retain many of its staff.

Cuban drug first

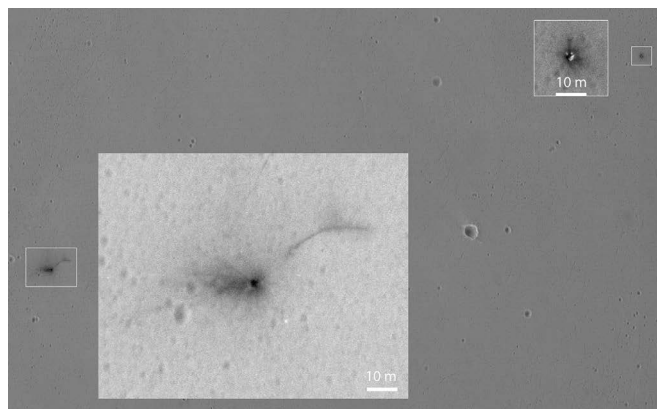
A cancer research centre in New York will host the first ever US clinical trial of a biotechnology developed in Cuba, state governor Andrew Cuomo announced on 26 October. The trial — a collaboration between Havana's Center of Molecular Immunology (CIM) and the Roswell Park Cancer Institute

NASA/JPL in Buffalo — will test the CIM's therapeutic vaccine CIMAvax-EGF in people with lung cancer. The vaccine is already approved for use in at least five countries, and researchers at both institutions think that it could eventually be used to prevent lung cancer in people at risk. The trial is a sign of thawing relations between Cuba and the United States, following the announcement last month of a US Treasury policy that authorizes US scientists to collaborate more freely with their Cuban counterparts.

SPACE

Mars crash site

NASA has released more-detailed images of the site on Mars where the European Space Agency's Schiaparelli lander met its end last month. Images taken by NASA's Mars Reconnaissance Orbiter on 25 October show three crash sites about 1.5 kilometres from each other (**pictured**), and suggest that a shallow crater was created by the impact. Initial indications suggest that a computing error during Schiaparelli's six-minute landing manoeuvre may have caused the craft to believe it was at a lower altitude than it really was and to jettison its parachute



too early. The craft, part of a European–Russian mission, was intended to test landing technology for a future Mars mission.

CLIMATE CHANGE

Sulfur phase-out

Ships will be banned from using high-sulfur fuel starting in 2020. The 171 member states of the International Maritime Organization, which regulates international shipping, have agreed to limit the pollutant to 0.5% in fuels by 2020, compared with an average of 3.5% today. The decision, made on 27 October at a meeting in London, will reduce sulfur oxide emissions and, the organization hopes, “have a significant beneficial impact on the environment and on human health”. But some environmental groups have criticized the deal for

failing to tackle carbon dioxide emissions. The shipping industry, a growing source of greenhouse-gas emissions, is not covered by global climate agreements such as the 2015 Paris deal.

POLITICS

Turkish rectorors

Under state-of-emergency provisions declared after an attempted military coup in July, Turkish President Recep Tayyip Erdoğan issued a decree on 29 October giving him the power to directly appoint state-university rectorors without input from the universities themselves. Previously, universities held elections to create shortlists of candidate rectorors, from which the government selected three to propose to the president. The decree also dismisses more than 10,000 civil servants,

COMING UP

3 NOVEMBER

China launches the next-generation Long March 5 rocket on its inaugural flight, from Hainan Island.

8–9 NOVEMBER

Leading researchers explain how their advances break down barriers in science and society at this year's Falling Walls conference in Berlin.

www.falling-walls.com

including 1,267 university academics, on suspicion of having links with terrorist groups. The dismissed include 24 scholars who signed an ‘Academics for Peace’ petition in January that called for an end to violence between government forces and Kurdish separatists. Critics say that people with no connection to terrorist groups are being targeted. An earlier decree in September dismissed more than 2,300 university academics, including 44 who had signed the peace petition.

PEOPLE

Prion pioneer dies

Susan Lindquist, an influential molecular biologist, died of cancer on 27 October, aged 67. Lindquist did pioneering work on heat-shock proteins, which can fix misfolded proteins. She was also recognized for her work on infectious proteins, known as prions, in yeast that cause Creutzfeldt–Jakob disease and other neurological disorders in humans. She spent 23 years at the University of Chicago in Illinois. In 2001, Lindquist joined the Whitehead Institute in Cambridge, Massachusetts, where she served as its first female director.

► NATURE.COM

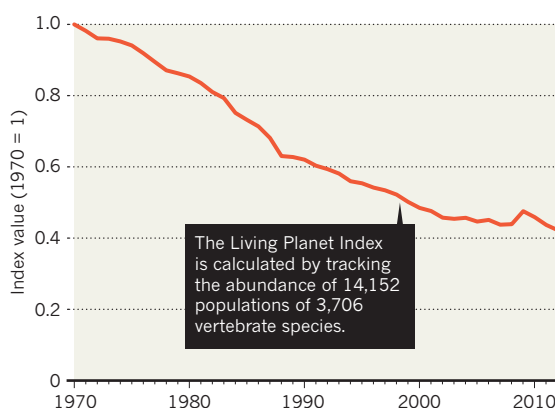
For daily news updates see:
www.nature.com/news

TREND WATCH

Wild populations of mammals, birds, amphibians, fish and other vertebrates declined by 58% between 1970 and 2012, according to the *Living Planet Report 2016*, published on 27 October. Freshwater populations, which fell by 81%, are thought to be faring worse than terrestrial ones. Habitat loss is the main threat, with overexploitation and human-induced climate change also major culprits. If the trend continues, by 2020 the world will have lost two-thirds of its vertebrate biodiversity, says the report.

VERTEBRATE NUMBERS IN DRASTIC DECLINE

Earth's vertebrate populations fell by 58% between 1970 and 2012, with human activities much to blame.



NEWS IN FOCUS

BIOTECHNOLOGY Synthetic biologists swap postdoc positions for start-ups **p.14**

HELIOPHYSICS Cosmic rays may threaten US space-weather satellite **p.15**

MENTAL HEALTH Moves to make psychiatry more mathematical **p.18**

ENERGY Can Africa leapfrog dirty fuels and go straight to renewables? **p.20**

PAUL NICKLEN/NATIONAL GEOGRAPHIC/GETTY



The Ross Sea in the Southern Ocean is one of the least-altered ecosystems on Earth and teems with a diverse array of marine life.

CONSERVATION

Giant ocean reserve is a go

International agreement to create world's largest marine protected area near Antarctica is hailed as a diplomatic breakthrough.

BY QUIRIN SCHIERMEIER

It is a milestone for ocean conservation and Russia's relationship with the rest of the world. After years of unsuccessful talks, 24 nations and the European Union agreed on 28 October to create the largest marine reserve on Earth, around twice the size of Texas, in the Southern Ocean off the coast of Antarctica.

The international deal sets aside 1.55 million square kilometres of the Ross Sea, a deep Antarctic bay 3,500 kilometres south of New Zealand, from commercial fishing and mineral

exploitation. It is the first time that countries have joined together to protect a major chunk of the high seas — the areas of ocean that are largely unregulated because they do not fall under the jurisdiction of any one nation. The deal takes effect in December 2017.

Signed by members of the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR) amid cheering at a meeting in Hobart, Australia, the deal became possible because of assent from Russia, which had long blocked the agreement. "Russian support of any agreement is a very positive signal

in the current political situation," says Peter Jones, a specialist in marine environmental governance at University College London.

Scientists hope now to see an acceleration of efforts to protect marine areas, in particular other ecologically precious regions around Antarctica. The designated reserve is a "first dent into the notion that we can't do anything to protect the high seas", says Daniel Pauly, a marine biologist at the University of British Columbia in Vancouver, Canada, who has long sounded the alarm over the state of the world's oceans. ►

► CCAMLR members had discussed the Ross Sea reserve since the United States and New Zealand proposed it in 2012. Observers say that Russia's change of heart might have been the result of behind-the-scenes discussions on the issue in recent months between the US secretary of state, John Kerry, and his Russian counterpart, Sergey Lavrov.

The Ross Sea is relatively healthy, but fishing activity is increasing — and that has begun to affect stocks of the predatory Antarctic toothfish (*Dissostichus mawsoni*). Also in decline is the Antarctic krill (*Euphausia superba*), a shrimp-like crustacean and a key creature in the marine food web off Antarctica.

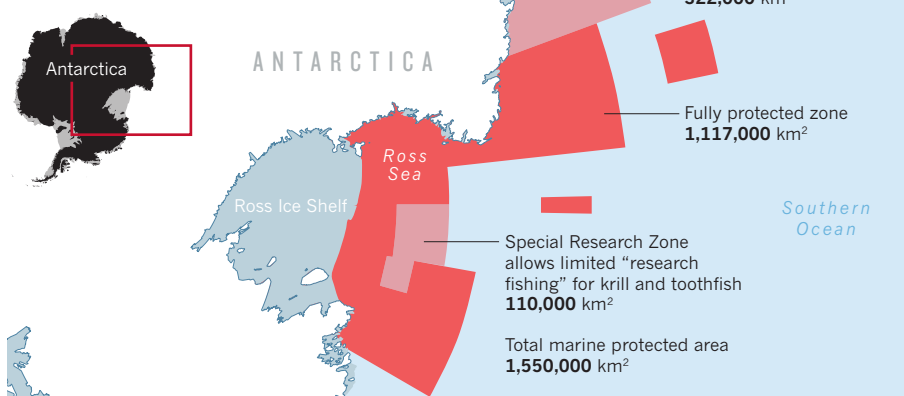
The deal includes some compromises. These might have been necessary to win the support of Russia, which operates a large fishing fleet in the region, says Jones.

Most of the reserve — 1.2 million km² — will be closed to all commercial marine activities. But a further 322,000-km² Krill Research Zone will allow controlled fishing, known as “research fishing”, and another 110,000 km² will be a Special Research Zone open for limited fishing of both krill and toothfish (‘see ‘Safeguarding the sea’). So although the total area of the marine reserve is bigger than the next-largest — Papahānaumokuākea Marine National Monument near Hawaii — the region that is completely restricted is slightly smaller.

And for now, a ‘sunset clause’ specifies that the designated zone will expire in 35 years, meaning that it does not fully qualify as a marine protected area (MPA) under the strict rules set by the International Union for Conservation of Nature. “We do regret this,” says Mike Walker, project director of the Antarctic Ocean Alliance, an environmental group in Washington DC.

SAFEGUARDING THE SEA

The newly created marine reserve in the Ross Sea near Antarctica includes different levels of protection.



“But we are confident that decision-makers will come to realize that the best way to conserve the ocean is to protect it forever.”

SCIENTIFIC PRAISE

On the whole, scientists reacted enthusiastically to the decision. The Ross Sea contains one of the least-altered ecosystems on Earth, says Kirsten Grorud-Colvert, a marine biologist at Oregon State University in Corvallis. But that ecosystem is vulnerable to human disturbance and the effects of climate change. “Setting aside an area free from fishing stresses in this marine reserve provides a reference point and a place for research to evaluate how systems respond to climate change, and to learn how to foster resilience,” she says.

Pauly adds, “It means we will protect one of the last parts of the world with a functioning natural ecosystem, with a complete array

of marine mammals, seabirds and other marine life.”

But others caution that ocean protection zones alone will not stop the decline in marine biodiversity, and that they do not provide a solution to overfishing because they may just move fishing to another spot. “If fishing is the problem, then they should reduce fishing pressure, not move it around,” says Ray Hilborn, a fisheries specialist at the University of Washington in Seattle. “Indeed, MPA might also stand for ‘Move Problems Elsewhere.’”

Next year, the CCAMLR will discuss further proposals to create protected zones of roughly similar size off the coast of East Antarctica and in the Weddell Sea. Chile and Argentina, meanwhile, are working on a proposal to protect the high seas surrounding the Antarctic Peninsula, the most rapidly warming part of the frozen continent. ■

BUSINESS

Young scientists gamble on biotech start-ups

Many are founding their own firms as venture capitalists show increased interest in science.

BY ERIKA CHECK HAYDEN

Vindication was three years coming for Ethan Perlstein. On 19 October, his California biotechnology company, Perlara, announced a deal with Novartis. The Swiss drug giant will test a compound that Perlara has identified as a possible treatment for a rare childhood disease, and will invest an undisclosed sum in the smaller firm.

Numerous biotech investors turned Perlstein away before he started Perlara in

San Francisco in 2014, because he wasn't the tenured professor that most venture capitalists saw as founder material. “They pretty much told me to take a hike,” he recalls.

But he persevered, and is now part of the vanguard of young biomedical scientists who have started companies instead of taking the conventional academic path and pursuing postdoctoral studies after their PhDs. Among the factors driving this change are an infusion of money into early-stage biotech investing, the emergence of biotech incubators and the

scarcity of academic jobs in science.

“We're starting to see a renaissance of investors embracing the idea that scientists can build businesses,” says Ryan Bethencourt, programme director of IndieBio, a biotech accelerator in San Francisco that began in 2014.

Previously, Bethencourt says, investors preferred to fund companies started by established professors who focused on the science, while investors installed a management team to take care of the business side. But that has changed as crucial technologies,

such as genetic sequencing, have become cheaper and lab work has become automated. The cost of starting biotech companies is falling, lowering the risk for investors to fund new science-based companies. IndieBio and Y Combinator — an information-technology incubator in Mountain View, California, that started accepting biotech companies in 2014 — provide funding and mentoring to entrepreneurs in exchange for shares in the companies.

FORK IN THE ROAD

Y Combinator, which provides US\$120,000 in seed funding per company, invested in Perlara this year; IndieBio, which provides \$250,000 per start-up, has funded 42 companies in a variety of fields. Last year, biotech firms in the United States and Europe raised \$3.5 billion in early-stage financing — more than in any previous year, according to the consultancy Ernst & Young. Much of this was from investors who have already made money in technology.

“Most of the venture guys I know want to change the world for the better,” says Dan Widmaier, co-founder and chief executive

of Bolt Threads in Emeryville, California, which uses genetic engineering to manufacture textiles. Widmaier went to work for the company three days after completing his PhD in 2010. “As they see it, being able to serve up an ad faster probably isn’t changing the world for the better as much as being able to solve climate change or cure disease.”

Conventional academic paths are also becoming less appealing. On average, young scientists earn their first US National Institutes of Health R01 grant — the bread-and-butter support for most biomedical scientists — at the age of 42. When Anitha Jayaprakash earned her genetics PhD from the Icahn School of Medicine at Mount Sinai in New York City in 2014, she saw scientists all around her stuck in postdocs. Many had no hope of finding their own tenure-track academic jobs — a phenomenon that Perlstein has dubbed the “postdocapocalypse”. “It gave me a very depressing feeling about the whole academic space,”

“Most of the venture guys I know what to change the world for the better.”

becoming less appealing. On average, young scientists earn their first US National Institutes of Health R01 grant — the bread-and-butter support for most biomedical scientists — at the age of 42. When Anitha Jayaprakash earned her genetics PhD from the Icahn School of Medicine at Mount Sinai in New York City in 2014, she saw scientists all around her stuck in postdocs. Many had no hope of finding their own tenure-track academic jobs — a phenomenon that Perlstein has dubbed the “postdocapocalypse”. “It gave me a very depressing feeling about the whole academic space,”

says Jayaprakash. So she started Girihlet, a genetic-sequencing company in Berkeley, California, that has received funding from IndieBio and other investors.

Alexander Lorestani felt the same way when he left a joint graduate and medical-degree programme in 2015 to co-found Geltor in San Leandro, California, which makes a vegan alternative to animal gelatin. He and his co-founder are 29 and 30 years old, and felt ready to use science to serve humanity. “I couldn’t imagine waiting another five to ten years to dive into doing what I think of as my life’s work,” Lorestani says.

It’s not an easy road. Most young biotech firms fail. Widmaier says that he never expected Bolt Threads to raise \$90 million and last for 6 years. He says it has been rewarding to thrive long enough to be doing groundbreaking science — and to have a rare degree of independence. “Anywhere else, you join someone else’s vision for what a perfect workplace is,” he says. “The most valuable thing about building a company is that you get to build the place where you go to work every day.” ■

HELIOPHYSICS

Hiccups for US satellite

Cosmic rays may be inducing glitches in space-weather probe’s computer.

BY ALEXANDRA WITZE

A space-weather satellite that is supposed to alert Earth to incoming solar storms has temporarily dropped offline six times in the year since it became operational. The US craft’s onboard computer may be experiencing hiccups caused unexpectedly by Galactic cosmic rays.

The Deep Space Climate Observatory (DSCOVR) went out of action most recently on 30 October. In each case, it unexpectedly entered a ‘safe hold’, in which scientific data stopped flowing and engineers had to scramble to try to recover the spacecraft. In total, DSCOVR’s space-weather forecasting instruments have been offline for more than 42 hours since 28 October 2015, when the US National Oceanic and Atmospheric Administration (NOAA) took the spacecraft over from NASA, which built and launched it.

Each outage lasts for only a few hours, and the total downtime amounts to more than 0.5% of its time in space — well within NOAA’s requirement that the spacecraft operate at least 96% of the time. The 11 October outage did not significantly affect predictions of a minor geomagnetic storm that arrived a



Workers test solar arrays on the DSCOVR satellite, which is now used to monitor space weather.

few days later, says Robert Rutledge, head of the forecast office at NOAA’s Space Weather Prediction Center in Boulder, Colorado.

But the outages mean that DSCOVR could be offline when a major solar storm erupts, leaving Earth essentially blind to the incoming

onslaught. “Are they problematic? Yes,” says Douglas Biesecker, a solar physicist at the Boulder centre.

Other heliophysics spacecraft monitor solar eruptions, but DSCOVR delivers unique information from its location at ►

► the gravitationally stable L1 point, about 1.5 million kilometres from Earth in the direction of the Sun. The spacecraft's instruments measure the speed, magnetic field and other properties of the charged particles streaming off the Sun. Those data translate into better forecasts of what could happen when a solar storm hits Earth, such as disruptions to satellite electronics or fluctuations in electrical power grids.

DSCOVR is NOAA's main tool for forecasting space weather, but it began life as Triana, a NASA Earth-observing spacecraft built in the late 1990s to gaze constantly at the planet. A pet project of Al Gore, then the US vice-president, Triana was shelved in 2001, then repurposed in 2008 for space-weather needs. "It was never designed from the beginning to be a space-weather satellite," says Steven Clarke, head of NASA's heliophysics division in Washington DC.

The satellite launched on 11 February 2015 and experienced its first outage four months later, when its onboard computer spontaneously rebooted. On average, the safe holds happen every 74 days, but two came just 8 days apart. They are not correlated with solar storms.

A NASA internal review board convened to study the problem could not definitively pinpoint the cause, but concluded that it was most likely to be Galactic cosmic rays randomly striking the spacecraft, causing high-energy ionization that reboots the computer. The computer, which was built by NASA in 2000, contains a processor card that is similar to those flying aboard many other missions and is meant to withstand the radiation hazards of deep space.

NOAA does have a back-up data stream, from the Advanced Composition Explorer (ACE) spacecraft that is also orbiting the L1 point. That was the primary source of solar-wind data until NOAA forecasters switched to DSCOVR in July. But ACE is 19 years old, and intense solar storms can swamp its forecasting instruments.

NOAA has requested an extra US\$1.5 million from Congress to improve how it handles DSCOVR data, including its responses to the outages. The satellite is supposed to last until 2022, when a follow-up mission is slated for launch. Historically, NOAA has cobbled together its space-weather observations where and when it could, but the US government is starting to demand a more coherent approach. On 13 October, President Barack Obama signed an executive order that, among other things, requires NOAA to "ensure the continuous improvement of operational space weather services". ■



Crop yields and water efficiency could be improved with the use of better gene-editing techniques.

BOTANY

A better way to hack plant DNA

As gene editing opens doors, crop researchers are hamstrung by the need for more-modern tools.

BY HEIDI LEDFORD

When crop engineers from around the world gathered in London in late October, their research goals were ambitious: to make rice that uses water more efficiently, cereals that need less fertilizer and uberproductive cassava powered by turbocharged photosynthesis.

The 150 attendees of the Crop Engineering Consortium Workshop were awash with ideas and brimming with molecular gadgets. Thanks to advances in synthetic biology and automation, several projects boasted more than 1,000 engineered genes and other molecular tools, ready to test in a researcher's crop of choice. But that is where they often hit a wall. Outdated methods for generating plants with customized genomes — a process called transformation — are cumbersome, unreliable and time-consuming.

Asked what hurdles remain for the field, plant developmental biologist Giles Oldroyd of the John Innes Centre in Norwich, UK, had a ready answer: "The big thing would be to

improve plant transformation," he said.

"What we're all facing is this delivery problem," says Dan Voytas, a plant biologist at the University of Minnesota in Saint Paul. "We have powerful reagents, but how do you get them into the cells?"

At issue is the decades-old problem that it is difficult to modify plant genomes and then regenerate a whole plant from a few transformed cells. Genome-editing techniques such as CRISPR-Cas9 hold out the promise of sophisticated crop engineering that would once have been unthinkable — making it all the more frustrating when researchers run up against an old roadblock.

On 28 September, the US National Science Foundation (NSF) recognized this frustration by announcing that it would fund research into better transformation methods. That focus is one of four in a new plant-genome research programme that will receive a total of US\$15 million.

"Everybody agrees that it really is the bottleneck for genome engineering," says Neal Stewart, a plant biologist at the University of

AFP/GETTY

Tennessee, Knoxville, who co-organized an NSF workshop about plant transformation last November. “And I think there’s enough interest now in trying to come up with ways to fix the problem for major crops.”

OBSTINATE CROPS

Some plants, such as the diminutive thale cress (*Arabidopsis thaliana*), the ‘lab rat’ of plants, are easily transformed using a bacterium that can add genes to plant genomes. Researchers insert the genes they want to test into the bacterium (*Agrobacterium tumefaciens*), and then coax the microbe to infect the reproductive cells of the plant. When the plant then produces offspring, some of them express the new genes.

But this does not work for many crops, and use of *Agrobacterium* triggers extra scrutiny from government agencies such as the US Department of Agriculture because it is considered a plant pest. As an alternative, researchers can use ‘gene guns’ that fire DNA-coated gold beads into plant cells. Those cells are then bathed in growth hormones and coaxed to regenerate a full plant. Some plants, such as maize (corn), readily bend to this treatment. Others, such as wheat and sorghum, do not.

For recalcitrant crops, it can take months of painstaking cell-culture work — optimizing growth conditions and hormone concentrations — to regenerate the full plant. The conditions needed for success vary not only from crop to crop, but also between plants of the same species.

Plant-transformation experts are a rare breed, says Joyce van Eck, one such specialist at Cornell University in Ithaca, New York. “There’s a lot of art in what we do,” she said at the London workshop. “It’s difficult to find people with that training.”

Add to that a dearth of funding for new methods, and researchers are left having to rely on decades-old techniques.

A BETTER WAY

But that could change as the hunt for alternatives heats up. Stewart and his collaborators have developed a robot that performs an established technique called protoplast transformation faster and more accurately than is possible by hand. The method uses enzymes to digest the cell wall, making it easier for researchers to introduce new genes. The problem of regenerating the whole plant, however, remains. Researchers used a similar approach,

without robots, to perform CRISPR–Cas9 gene editing in a variety of plants, including lettuce and rice.

The cell-culture steps are still difficult. Stewart says that one person in his lab laboured unsuccessfully for two years to transform a tall grass that he uses for biofuel research. But the declining cost of enzymes allows researchers to perform more experiments, and the robotics improve throughput. Stewart is so enamoured with his creation that he has composed a song for it. “It’s our baby right now,” he says.

Others, such as Fredy Altpeter of the University of Florida in Gainesville, are hunting for a suite of genes that, when switched on or off, would make plant cells more amenable to transformation and regeneration from culture. “I think it will lead to much broader application of this technology, and will enable people who are not experts in cell culture to make those improvements,” he says.

But researchers can’t afford to wait for those developments, says Oldroyd. His project, which aims to develop cereals that use nitrogen from the soil more efficiently, will plough through tests of hundreds of transgenes using the old, cumbersome methods. “We just have to be patient,” he says. ■

INFECTIOUS DISEASE

Infected mosquitoes fight Zika

South America hosts largest trials yet of *Wolbachia*-infected insects to combat viruses.

BY EWEN CALLAWAY

Two South American metropolises are enlisting bacterium-infected mosquitoes to fight Zika. The effort is the world’s biggest test yet of an unconventional but promising approach to quell mosquito-borne diseases.

Mosquitoes that carry *Wolbachia* bacteria — which hinder the insects’ ability to transmit Zika, dengue and other viruses — will be widely released in Rio de Janeiro, Brazil, and Medellín, Colombia, over the next two years, scientists announced on 26 October. The deployments will reach around 2.5 million people in each city. “This really has the potential to be a game changer in terms of vector control — the biggest thing since DDT,” says Philip McCall, a medical entomologist who studies mosquito control at the Liverpool School of Tropical Medicine, UK.

Small numbers of *Wolbachia*-infected mosquitoes have already been released in both Rio de Janeiro and Medellín. But large biomedical funders have now announced US\$18 million to scale up the efforts. “We really want to deploy quite quickly in large sections of these

cities,” says Scott O’Neill, a microbiologist at Monash University in Melbourne, Australia, and head of the Eliminate Dengue Program, which is leading the mosquito releases. Footing the bill are the Bill & Melinda Gates Foundation in Seattle, Washington, the London-based Wellcome Trust and the US and UK governments. Brazil’s government is chipping in with an extra \$3.7 million, O’Neill says.

VIRUS BLOCKERS

Wolbachia pipiensis plagues some 60% of insect species worldwide — but doesn’t naturally infect *Aedes aegypti* mosquitoes, the species that transmits Zika, dengue and numerous other viruses. The bacteria can hinder their hosts’ fertility and influence the sex of offspring. They can also block viruses from reproducing in infected fruit flies and mosquitoes, as O’Neill and his colleagues discovered in the late 1990s. The team later developed laboratory populations of infected *A. aegypti*.

When tens of thousands of these mosquitoes were released near the small city of Cairns in northern Australia in 2011, the bacteria spread rapidly among local *A. aegypti* mosquitoes; 90% of mosquitoes in a targeted area were infected within weeks. Tests in Indonesia and

Vietnam found similar success. It’s not yet clear whether the strategy also reduces rates of dengue infections in humans, but O’Neill’s team has begun a trial in Yogyakarta, Indonesia, to find out.

The Eliminate Dengue team started releasing *Wolbachia*-infected mosquitoes in two Rio de Janeiro neighbourhoods in 2014, and in a suburb of Medellín in 2015. The

bacteria block the replication of Zika and chikungunya virus (which caused widespread outbreaks in Latin America and the Caribbean in 2013–14). O’Neill’s team hopes that the scaled-up deployments can combat those diseases, as well as dengue, which infected



Aedes aegypti mosquitoes spread Zika, dengue and other viruses.

ISTOCK/GETTY

► an estimated 1.6 million people in Brazil last year. The researchers plan to release mosquitoes in waves, survey the insects for infection and track the incidence of disease in areas with and without infected mosquitoes.

Other scientists are also testing *Wolbachia* to control mosquitoes. In Singapore, officials plan to release male *A. aegypti* mosquitoes infected with a strain of *Wolbachia* that renders their offspring infertile. A US biotechnology company is seeking approval to use a similar approach to combat related Asian tiger mosquitoes (*Aedes albopictus*), which carry dengue and chikungunya. In Guangzhou, China, scientists are releasing *A. albopictus* mosquitoes infected with *Wolbachia* each week, in large-scale field trials. And researchers in French Polynesia are trying the same strategy on another species of tiger mosquito.

MORE EVIDENCE NEEDED

Wolbachia has an impressive ability to surge through wild mosquito populations, says McCall, but proving that this limits human infections will be critical before the approach can find widespread use. If *Wolbachia* is to make a dent in mosquito-borne diseases, the technique will also have to be cost-effective and long-lasting, he adds. “If it works, it will be truly remarkable, but it has to still be working in ten years.”

Another hurdle facing the tests in Rio de Janeiro and Medellín is the size of the cities — especially Rio, with its densely populated and hard-to-access *favelas*, says McCall. But if *Wolbachia* can combat Zika, dengue and chikungunya in such environments, “there is a very strong case for doing it for a range of other large cities”, says Mike Turner, acting director of science at the Wellcome Trust. Widespread deployment of *Wolbachia*-infected mosquitoes would probably also depend on endorsement from the World Health Organization, he adds.

Public support could make or break the *Wolbachia* approach, says O'Neill, whose team spent years engaging with communities in Australia before deploying mosquitoes there. *Wolbachia* is already widespread among insects and it cannot infect humans, he notes. In Australia, researchers recruited schoolchildren, whom they dubbed *Wolbachia* warriors, to rear the eggs at home, learn about their development and then release the mosquitoes.

In Colombia, the Eliminate Dengue team has worked with “Casa *Wolbachia*” families to help with the release of mosquitoes, and even written salsa songs about the bacteria, says co-principal investigator Jorge Osorio, a pathobiologist at the University of Wisconsin–Madison. “We have communities asking us to spread more mosquitoes,” he says. ■

Q&A Joshua Gordon

Psychiatry needs more mathematics

The US National Institute of Mental Health (NIMH) has a new director. On 12 September, psychiatrist Joshua Gordon took the reins at the institute, which has a budget of US\$1.5 billion. He previously researched how genes predispose people to psychiatric illnesses by acting on neural circuits, at Columbia University in New York City. His predecessor, Thomas Insel, left the NIMH to join Verily Life Sciences, a start-up owned by Google's parent company Alphabet, in 2015. Gordon says that his priorities at the NIMH will include “low-hanging clinical fruit, neural circuits and mathematics — lots of mathematics”, and explains to Nature what that means.



Second, I will be thinking about how NIMH research can be structured to give payouts in the short, medium and long terms.

How has neuroscience changed since you completed your medical residency in 2001?

The advent of incredibly powerful tools to observe and alter activity in a subset of neurons, such as optogenetics, has been transformational. It is allowing us to get at questions of how neural circuits produce behaviour — a research approach that may soon generate new treatments for psychiatric disorders.

Which of the recent NIMH programmes do you find particularly exciting?

One is the Human Connectome Project. The project has scanned the brains of more than 1,000 healthy people to generate individual maps of their neural circuitry, the ‘wiring’ in their brains that accounts for their particular personalities. At the NIMH, we have created standardized databases, designed by the scientific community, to store this information. The Human Connectome Project is going to be a tremendous resource for the field — maybe not quite as impactful as the Human Genome Project, but on that scale, I think.

A clinical programme that deserves as much attention, but perhaps doesn't get it, is the Coordinated Specialty Care project for individuals facing their first psychotic episode. Some small studies have shown that coordinating different clinical and social-support programmes helps individuals to cope better.

Is this an example of ‘low-hanging fruit’?

Yes. We are now looking for similarly significant clinical problems where good, evidence-based interventions exist but are not widely adopted. For example, we have a range of screening tools that we think can help reduce the suicide rate, which has been rising in the United States. It could be advantageous to incorporate universal suicidality screening as a matter of routine into all emergency rooms.

What do you plan to achieve in your first year?

I won't be doing anything radical. I am just going to listen to and learn from all the stakeholders — the scientific community, the public, consumer advocacy groups and other government offices. But I can say two general things.

In the past 20 years, my two predecessors, Steve Hyman [now director of the Stanley Center for Psychiatric Research at the Broad Institute in Cambridge, Massachusetts] and Tom Insel, embedded into the NIMH the idea that psychiatric disorders are disorders of the brain, and to make progress in treating them we really have to understand the brain. I will absolutely continue this legacy. This does not mean we are ignoring the important roles of the environment and social interactions in mental health — we know they have a fundamental impact. But that impact is on the brain.

What about medium-term payouts?

Neural circuits could be delivering treatments in 10 or 15 years. We don't yet know exactly which circuits we would want to modify to treat psychiatric disorders in humans. But now is the time to start thinking about which tools we are going to need to make this translational step possible, and invest in them.

Most work on neural circuits has been done in genetically modified mice, where it is relatively easy to control the activity of a few very specific cells in a particular brain area using tools such as optogenetics. We'll need safe methods for humans. Should we be thinking in terms of viruses that can be directed to, and change activity in, specific neurons? Or should we be thinking of ways to stimulate or inhibit these cells indirectly, using transcranial magnetic stimulation or deep-brain stimulation,?

And the long term?

The really transformative treatments that are going to change mental-health care in the long term will depend on us learning how the brain works as a whole. We are all tempted to reduce the huge complexity of the brain into understandable chunks. But to appreciate and exploit that complexity, we will need to be able to integrate everything we know, from molecular biology to behaviour, into our models of how the brain works. That requires serious math.

How does the structure of a neuron affect its integration into a circuit? How does that circuit affect the neural system that it fits into? How does the dynamic activity in these neural systems drive behaviour? Fully characterizing each of these levels and then integrating them across scales requires a level of mathematical rigour that most of us, including myself, have not really brought to bear on the problem.

Isn't the mathematics going to get very difficult for neuroscientists?

It's not so difficult — I'm not saying that we are going to need string theorists! It's just a question of appropriate training in math for students. In the future, I hope that every experimentalist will also be a theoretician. But at this stage we need to encourage experimental neurobiologists to form long-term interdisciplinary collaborations with theoreticians, mathematicians or physicists. We need to inject more math into every level of the NIMH portfolio. Math can also have a short-term impact in psychiatry for things such as predicting individual responses to drugs and improving precision medicine more generally.

Are non-human primates still necessary in neuroscience research?

Most of our knowledge about the brain has been gained in mice. It is hard for me to believe

that we'll really be able to translate the knowledge that we have won in mice into the design of new treatments for humans without going through an intermediate species with an elaborated prefrontal cortex and a large brain. So unfortunately, yes, I think we do still need to use non-human primates. We need to do so judiciously, though — the welfare of animals is fundamental, and we need to minimize the numbers of all of the animals that we use.

In a move to a circuits-based approach, the NIMH introduced the Research Domain Criteria (RDoC), which encourages clinical researchers to investigate specific behaviours rather than broad diagnoses. It is widely disliked — will you be maintaining it?

Clinical neuroscience has typically tried to identify the neurobiology that underlies diagnoses [such as depression]. That hasn't got us very far. Maybe if we instead try to understand the neurobiology underlying the various domains of behaviour [such as apathy], we'll get better insight. I see RDoC as something potentially very valuable, something I am likely to keep — although it may need a few tweaks to extract the most value out of it. ■

INTERVIEW BY ALISON ABBOTT

This interview has been edited for length and clarity. See go.nature.com/2f88hif for a longer version.



Can wind and solar fuel Africa's future?

With prices for renewables dropping, many countries in Africa might leap past dirty forms of energy towards a cleaner future.

BY ERICA GIES

At the threshold of the Sahara Desert near Ouarzazate, Morocco, some 500,000 parabolic mirrors run in neat rows across a valley, moving slowly in unison as the Sun sweeps overhead. This US\$660-million solar-energy facility opened in February and will soon have company. Morocco has committed to generating 42% of its electricity from renewable sources by 2020.

Across Africa, several nations are moving aggressively to develop their solar and wind capacity. The momentum has some experts wondering whether large parts of the continent can vault into a clean future, bypassing some of the environmentally destructive practices that have plagued the United States, Europe and China, among other places.

“African nations do not have to lock into developing high-carbon old technologies,” wrote Kofi Annan, former secretary-general of the

United Nations, in a report last year¹. “We can expand our power generation and achieve universal access to energy by leapfrogging into new technologies that are transforming energy systems across the world.”

That’s an intoxicating message, not just for Africans but for the entire world, because electricity demand on the continent is exploding. Africa’s population is booming faster than anywhere in the world: it is expected to almost quadruple by 2100. More than half of the 1.2 billion people living there today lack electricity, but may get it soon. If much of that power were to come from coal, oil and natural gas, it could kill international efforts to slow the pace of global warming. But a greener path is possible because many African nations are just starting to build up much of their energy infrastructure and have not yet committed to dirtier technology.

Several factors are fuelling the push for renewables in Africa. More

MAINSTREAM RENEWABLE POWER

Jeffreys Bay Wind Farm in South Africa generates enough energy to power 100,000 homes there.

than one-third of the continent's nations get the bulk of their power from hydroelectric plants, and droughts in the past few years have made that supply unreliable. Countries that rely primarily on fossil fuels have been troubled by price volatility and increasing regulations. At the same time, the cost of renewable technology has been dropping dramatically. And researchers are finding that there is more potential solar and wind power on the continent than previously thought — as much as 3,700 times the current total consumption of electricity.

This has all led to a surging interest in green power. Researchers are mapping the best places for renewable-energy projects. Forward-looking companies are investing in solar and wind farms. And governments are teaming up with international-development agencies to make the arena more attractive to private firms.

Yet this may not be enough to propel Africa to a clean, electrified future. Planners need more data to find the best sites for renewable-energy projects. Developers are wary about pouring money into many countries, especially those with a history of corruption and governmental problems. And nations will need tens of billions of dollars to strengthen the energy infrastructure.

Still, green ambitions in Africa are higher now than ever before. Eddie O'Connor, chief executive of developer Mainstream Renewable Power in Dublin, sees great potential for renewable energy in Africa. His company is building solar- and wind-energy facilities there and he calls it "an unparalleled business opportunity for entrepreneurs".

POWER PROBLEMS

Power outages are a common problem in many African nations, but Zambia has suffered more than most in the past year. It endured a string of frequent and long-lasting blackouts that crippled the economy. Pumps could not supply clean water to the capital, Lusaka, and industries had to slash production, leading to massive job lay-offs.

The source of Zambia's energy woes is the worst drought in southern Africa in 35 years. The nation gets nearly 100% of its electricity from hydropower, mostly from three large dams, where water levels have plummeted. Nearby Zimbabwe, South Africa and Botswana have also had to curtail electricity production. And water shortages might get worse. Projections suggest that the warming climate could reduce rainfall in southern Africa even further in the second half of the twenty-first century.

Renewable energy could help to fill the gap, because wind and solar projects can be built much more quickly than hydropower, nuclear or fossil-fuel plants. And green-power installations can be expanded piecemeal as demand increases.

Egypt, Ethiopia, Kenya, Morocco and South Africa are leading the charge to build up renewable power, but one of the biggest barriers is insufficient data. Most existing maps of wind and solar resources in Africa do not contain enough detailed information to allow companies to select sites for projects, says Grace Wu, an energy researcher at the University of California, Berkeley. She co-authored a report² on planning renewable-energy zones in 21 African countries, a joint project by the Lawrence Berkeley National Laboratory (LBNL) in California and the International Renewable Energy Agency (IRENA) in Abu Dhabi. The study is the most comprehensive mapping effort so far for most of those countries, says Wu. It weighs the amount of solar and wind energy in the nations, along with factors such as whether power projects would be close to transmission infrastructure and customers, and whether they would cause social or environmental harm. "The IRENA-LBNL study is the only one that has applied a consistent methodology across a large region of Africa," says Wu. High-resolution measurements of wind and solar resources have typically been done by government researchers or companies, which

kept tight control of their data. The Berkeley team used a combination of satellite and ground measurements purchased from Vaisala, an environmental monitoring company based in Finland that has since made those data publicly available through IRENA's Global Atlas for Renewable Energy. The team also incorporated geospatial data — the locations of roads, towns, existing power lines and other factors — that could influence decisions about where to put energy projects. "If there's a forest, you don't want to cut it down and put a solar plant there," says co-author Ranjit Deshmukh, also an energy researcher at Berkeley.

The amount of green energy that could be harvested in Africa is absolutely massive, according to another IRENA report³, which synthesized 6 regional studies and found potential for 300 million megawatts of solar photovoltaic power and more than 250 million megawatts of wind (see 'Power aplenty'). By contrast, the total installed generating capacity — the amount of electricity the entire continent could produce if all power plants were running at full tilt — was just 150,000 megawatts at the end of 2015. Solar and wind power accounted for only 3.6% of that.

The estimate of wind resources came as a surprise, says Oliver Knight, a senior energy specialist for the World Bank's Energy Sector Management Assistance Program in Washington DC. Although people have long been aware of Africa's solar potential, he says, as of about a decade ago, few local decision-makers recognized the strength of the wind. "People would have told you there isn't any wind in regions such as East Africa."

The World Bank is doing its own studies, which will assess wind speeds and solar radiation at least every 10 minutes at selected sites across target countries. It will ask governments to add their own geospatial data, and will combine all the information into a user-friendly format that is freely available and doesn't require advanced technical knowledge, says Knight. "It should be possible for a mid-level civil servant in a developing country to get online and actually start playing with this."

SOUTH AFRICA LEADS

In the semi-arid Karoo region of South Africa, a constellation of bright white wind turbines rises 150 metres above the rolling grassland. Mainstream Renewable Power brought this project online in July, 17 months after starting construction. The 35 turbines add 80 megawatts to South Africa's supply, enough to power about 70,000 homes there.

The Noupoort Wind Farm is just one of about 100 wind and solar projects that South Africa has developed in the past 4 years, as prices fell below that of coal and construction lagged on two new massive coal plants. South Africa is primed to move quickly to expand renewable energy, in part thanks to its investment in data.

Environmental scientist Lydia Cape works for the Council for Scientific and Industrial Research, a national lab in Stellenbosch. She and her team have created planning maps for large-scale wind and solar development and grid expansion. Starting with data on the energy resources, they assessed possible development sites for many types of socio-economic and environmental impact, including proximity to electricity demand, economic benefits and effects on biodiversity.

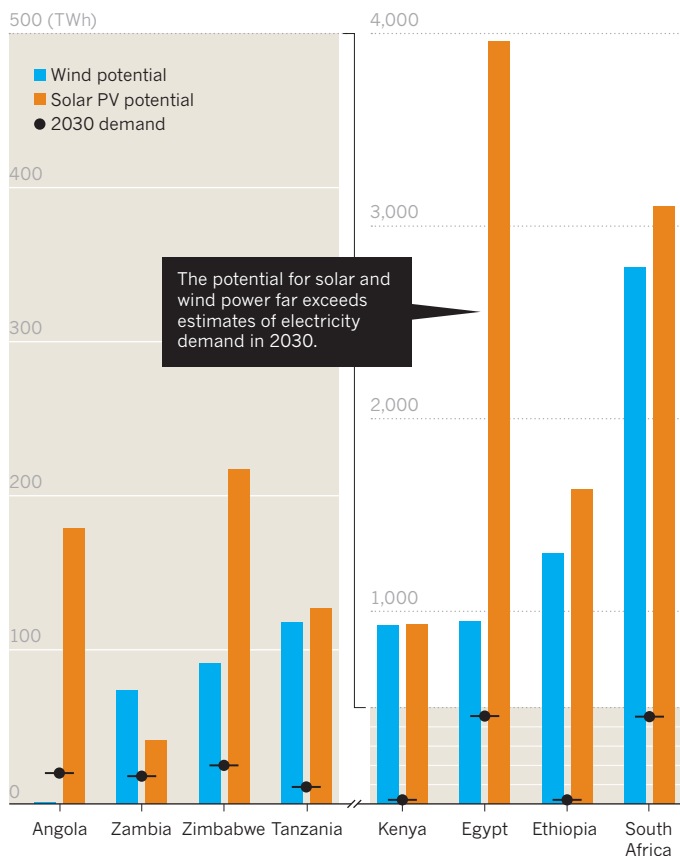
The South African government accepted the team's recommendations and designated eight Renewable Energy Development Zones that are close to consumers and to transmission infrastructure — and where power projects will cause the least harm to people and ecosystems. They total "about 80,000 square kilometres, the size of Ireland or Scotland, roughly", says Cape. The areas have been given streamlined environmental authorization for renewable projects and transmission corridors, she says.

But for African nations to go green in a big way, they will need a huge influx of cash. Meeting sub-Saharan Africa's power needs will cost US\$40.8 billion a year, equivalent to 6.35% of Africa's gross domestic product, according to the World Bank. Existing public funding falls

"A decade ago, people would have told you there isn't any wind in regions such as East Africa."

POWER APLENTY

Studies of some African nations suggest that they could harvest vast amounts of power from wind turbines and solar photovoltaic (PV) projects.



far short, so attracting private investors is crucial. Yet many investors perceive African countries as risky, in part because agreements there require long and complex negotiations and capital costs are high. "It's a real challenge," says Daniel Kammen, a special envoy for energy for the US Department of State and an energy researcher at the University of California, Berkeley. "Many of these countries have not had the best credit ratings."

Elham Ibrahim, the African Union's commissioner for infrastructure and energy, advises countries to take steps to reassure private investors. Clear legislation supporting renewable energy is key, she says, along with a track record of enforcing commercial laws.

South Africa is setting a good example. In 2011, it established a transparent process for project bidding called the Renewable Energy Independent Power Producer Procurement Programme (REIPPPP). The programme has generated private investments of more than \$14 billion to develop 6,327 megawatts of wind and solar.

Mainstream Renewable Power has won contracts for six wind farms and two solar photovoltaic plants through REIPPPP. "This programme is purer than the driven snow," says O'Connor. "They publish their results. They give state guarantees. They don't delay you too much." Although the country's main electricity supplier has wavered in its support for renewables, the central government remains committed to the programme, he says. "I would describe the risks in South Africa as far less than the risks in England in investing in renewables."

For countries less immediately attractive to investors, the World Bank Group launched the Scaling Solar project in January 2015. This reduces risk to investors with a suite of guarantees, says Yasser Charafi, principal investment officer for African infrastructure with the International Finance Corporation (IFC) in Dakar, which is part of the World Bank Group. Through the Scaling Solar programme, the IFC offers low-priced

loans; the World Bank guarantees that governments will buy the power generated by the projects; and the group's Multilateral Investment Guarantee Agency offers political insurance in case of a war or civil unrest.

Zambia, the first country to have access to Scaling Solar, has won two solar projects that will together provide 73 megawatts. Senegal and Madagascar were next, with agreements to produce 200 and 40 megawatts, respectively. Ethiopia has just joined, and the IFC will give two further countries access to the programme soon; its target is to develop 1,000 megawatts in the first 5 years.

MAKING IT FLOW

That power won't be useful if it can't get to users. One of the big barriers to a clean-energy future in Africa is that the continent lacks robust electricity grids and transmission lines to move large amounts of power within countries and across regions.

But that gap also provides some opportunities. Without a lot of existing infrastructure and entrenched interests, countries there might be able to scale up renewable projects and manage electricity more nimbly than developed nations. That's what happened with the telephone industry: in the absence of much existing land-line infrastructure, African nations rapidly embraced mobile phones.

The future could look very different from today's electricity industry. Experts say that Africa is likely to have a blend of power-delivery options. Some consumers will get electricity from a grid, whereas people in rural areas and urban slums — where it is too remote or too expensive to connect to the grid — might end up with small-scale solar and wind installations and minigrids.

Still, grid-connected power is crucial for many city dwellers and for industrial development, says Ibrahim. And for renewables to become an important component of the energy landscape, the grid will need to be upgraded to handle fluctuations in solar and wind production. African nations can look to countries such as Germany and Denmark, which have pioneered ways to deal with the intermittent nature of renewable energy. One option is generating power with existing dams when solar and wind lag, and cutting hydropower when they are plentiful. Another technique shuttles electricity around the grid: for example, if solar drops off in one place, power generated by wind elsewhere can pick up the slack. A third strategy, called demand response, reduces electricity delivery to multiple customers by imperceptible amounts when demand is peaking.

These cutting-edge approaches require a smart grid and infrastructure that connects smaller grids in different regions so that they can share electricity. Africa has some of these 'regional interconnections', but they are incomplete. Four planned major transmission corridors will need at least 16,500 kilometres of new transmission lines, costing more than \$18 billion, says Ibrahim. Likewise, many countries' internal power grids are struggling to keep up.

That's part of what makes working in energy in Africa challenging. Prosper Amuquandoh is an inspector for the Ghana Energy Commission and the chief executive of Smart and Green Energy Group, an energy-management firm in Accra. In Ghana, he says, "there's a lot of generation coming online".

The country plans to trade electricity with its neighbours in a West African Power Pool, Amuquandoh says, but the current grid cannot handle large amounts of intermittent power. Despite the challenges, he brims with enthusiasm when he talks about the future: "The prospects are huge."

With prices of renewables falling, that kind of optimism is spreading across Africa. Electrifying the continent is a moral imperative for everyone, says Charafi. "We cannot just accept in the twenty-first century that hundreds of millions of people are left out." ■

Erica Gies is a freelance journalist in Victoria, British Columbia.

1. Africa Progress Panel *Power, People, Planet: Seizing Africa's Energy and Climate Opportunities* (Africa Progress Panel, 2015).
2. Wu, G. C., Deshmukh, R., Ndhlukula, K., Radojicic, T. & Reilly, J. *Renewable Energy Zones for the Africa Clean Energy Corridor* (IRENA/LBNL, 2015).
3. Miketa, A. & Saadi, N. *Africa Power Sector: Planning and Prospects for Renewable Energy* (IRENA, 2015).

COMMENT

GEOSCIENCE Earth scientists of different stripes must talk to, not past, each other **p.25**

CRYPTOGRAPHY Journey to the heart of the world's most mysterious manuscript **p.28**

HISTORY Modern zoos emerged through adaptation and survival **p.29**

DEVELOPMENT There is a north-south inequality in inequality studies **p.31**



REIDAR HAHN/FERMILAB



The Dark Energy Survey at the Cerro Tololo Inter-American Observatory in Chile is mapping the large-scale structure of the Universe traced by galaxies.

Good data are not enough

A vibrant scientific culture encourages many interpretations of evidence, argues **Avi Loeb**.

This summer, I visited the Mayan city of Chichén Itzá in the Yucatán Peninsula, Mexico. It has an ancient observatory where priest-astronomers made detailed astronomical observations around AD 600–1200. The ruins — stepped pyramids, temples, columned arcades and other stone structures — reveal that astronomy was at the heart of this sophisticated society.

The Mayans accurately tracked changes in the positions and relative brightness of the Sun, Moon, planets and stars. They documented their astronomical data in folding books called codices, with many more quantitative details than other civilizations

at the time. The priest-astronomers used observations and advanced mathematical calculations to predict eclipses, and devised a 365-day solar calendar that was off by just one month every 100 years.

So why, I wondered, didn't the Mayans go further and infer aspects of our modern understanding of astronomy? They determined the orbital periods of Venus, Mars and Mercury around the Sun, but Earth was at the centre of their Universe.

I came to appreciate how limiting prevailing world views can be. Just as geological and other evidence for the great age of Earth was rejected before the nineteenth century as

being hard to square with biblical history, the Mayans used their fine data to support a mythological culture of astrology. They correlated the periodic motions of celestial objects with human history and, rather than seeking a physical explanation for their astronomical data, they used it to initiate wars or rituals such as human sacrifice.

Have we learned our lesson, or is today's science similarly trapped by cultural and societal forces? Most research funding is allocated assuming that the highest-quality data will inevitably deliver useful scientific interpretation and theoretical concepts, which can be tested and refined by future ▶

► data. The astronomy division of the US National Science Foundation, for example, devotes most of its funds to major facilities and large surveys, which are performed by big teams to collect better data within mainstream paradigms. Fields from particle physics to genomics do the same.

The consequences of a closed scientific culture are wasted resources and misguided ‘progress’ — witness the dead end that was Soviet evolutionary biology. To truly move forward, free thought must be encouraged outside the mainstream. Multiple interpretations of existing data and alternative motivations for collecting new data must be supported.

BLINKERED VIEW

Mayan cosmologists had high social status. They got generous support because of their promises to forecast the future. Cosmologists today collect vast amounts of exquisite data in surveys of large parts of the sky, costing billions of dollars.

Surveys of the large-scale structure of the Universe traced by galaxies include the US Baryon Oscillation Spectroscopic Survey and the international Dark Energy Survey, as well as forthcoming facilities such as the US Dark Energy Spectroscopic Instrument, the European Space Agency’s (ESA’s) Euclid mission, NASA’s Wide Field Infrared Survey Telescope and the Large Synoptic Survey Telescope in Chile. Others mapping the primordial seeds of these structures as traced by cosmic microwave background radiation include ESA’s Planck satellite, the US South Pole Telescope and international collaborations such as the Atacama Cosmology Telescope and the Simons Array.

Such projects have a narrow aim — pinning down the parameters of one theoretical model. The model comprises an expanding Universe composed of dark matter, dark energy and normal matter (from which stars, planets and people are made), with initial conditions dictated by an early phase of rapid expansion called cosmic inflation. The data are reduced to a few numbers. Surprises in the rest are tossed away.

I noticed this bias recently while assessing a PhD thesis. The student was asked to test whether a data set from a large cosmological survey was in line with the standard cosmological model. But when a discrepancy was found, the student’s goal shifted to explaining why the data set was incomplete. In such a culture, the current model can never be ruled out, even though everyone knows that its major constituents (dark matter, dark energy and inflation) are not understood at a fundamental level.

Instead, observers should present results in a theory-neutral way. Observations should not converge on one model but aim to find anomalies that carry clues about the nature of

dark matter, dark energy or initial conditions of the Universe. Further observations should be motivated by testing unconventional interpretations of those anomalies (such as exotic forms of dark matter or modified theories of gravity). Vast data sets may contain evidence for unusual behaviour that was unanticipated when the projects were conceived. If all results are expected and planned for, babies may be thrown out with the bathwater.

BLINDED BY BEAUTY

How each culture views the Universe is guided by its beliefs in, for example, mathematical beauty or the structure of reality. If these ideas are deeply rooted, people tend to interpret all data as supportive of them — adding parameters or performing mathematical gymnastics to force the fit. Recall how the belief that the Sun moves around Earth led to the mathematically beautiful (and incorrect) theory of epicycles advocated by the ancient Greek philosopher Ptolemy.

Similarly, modern cosmology is augmented by unsubstantiated, mathematically sophisticated ideas — of the multiverse, anthropic reasoning and string theory. The multiverse idea postulates the existence of numerous other regions of space-time, to which we have no access and in which the cosmological parameters have different values.

The anthropic argument is then often applied. It holds that our own region has the parameters it does (including those of dark energy and dark matter) because other, more likely values would not have allowed life to develop near a star like the Sun in a galaxy such as the Milky Way^{1–3}. An overlooked problem with this argument is that, according to one analysis⁴, life is 1,000 times more likely to exist 10 trillion years from now around stars that weigh one-tenth the mass of the Sun. This means that terrestrial life might be premature and not the most likely form of life, even in our own Universe⁴.

The anthropic argument, which as yet has no empirical support, suppresses

much-needed efforts to understand dark energy through an alternative theory that unifies quantum mechanics and gravity. The fact that we have not yet converged on such a theory is indicated by paradoxes in other areas of physics. For example, information contained in, say, an encyclopaedia is lost if it is swallowed by a black hole that ultimately evaporates into heat known as Hawking radiation. This contradicts a basic premise of quantum mechanics that information is preserved, and is known as the ‘information paradox’⁵. In addition, currently viable models of cosmic inflation require fine-tuning of the conditions of the Universe before and during inflation⁶.

Cultivating other approaches avoids stalling progress by investing only in chasing what might turn out to be ‘epicycles’. After all, the standard model of cosmology is merely a precise account of our ignorance: we do not understand the nature of inflation, dark matter or dark energy. The model has difficulties accounting for the luminous gas and stars that we can see in galaxies, while leaving invisible what we can easily calculate (dark matter and dark energy). This state of affairs is clearly unsatisfactory.

The tendency to establish large projects and firm up mainstream ideas is a signature of a mature scientific discipline. In such a culture, the low-hanging fruit has already been picked by small, versatile teams that are long gone. Critics argue that when funds are limited, the focus of research should be on coordinated approaches that are likely to produce results in a predictable way. This advocacy fails to appreciate that our mainstream paradigm might be heading in the wrong direction. The opportunity for making mistakes is much greater than for real breakthroughs, so as any venture capitalist knows, investing part of the portfolio in risky endeavours is necessary to gain substantial profits.

ALTERNATIVE PATHS

The only way to work out whether we are on the wrong path is to encourage competing interpretations of the known data.

I have been arguing for many years that funding agencies should promote the analysis of data for serendipitous purposes beyond major programmes and the mainstream dogma. The need for a change in course is even more pressing now. Empirical constraints on expected forms of dark matter (such as weakly interacting massive particles or supersymmetric partners to known particles) are getting tighter, and the hope of identifying testable consequences of string theory is receding. At a minimum, when funding is tight, a research frontier should maintain at least two ways of interpreting data so that new experiments will aim to select the correct one. This should apply to alternatives to inflation when dealing with new



The Mayan pyramid at Chichén Itzá in Mexico.

cosmological data, and to alternatives to cold dark matter when discrepancies are observed in the properties of dark-matter-dominated galaxies.

New funding streams should be established in other fields. The LIGO discovery of black-hole mergers should encourage a 'template-free' search for new sources of gravitational waves that were never imagined. The Kepler satellite's discovery that roughly one-quarter of all stars in the Galaxy host a habitable Earth-mass planet⁷ should lead to a renewed effort in the search for extraterrestrial life, including new methods for finding intelligent civilizations⁸. Indeed, a habitable planet was recently discovered⁹ around the nearest star to our Sun, Proxima Centauri, which could be probed with a future spacecraft (<http://breakthroughinitiatives.org/Concept/3>).

A healthy dialogue between different points of view should be fostered through multidisciplinary conferences that discuss conceptual issues, not just experimental results and phenomenology. A diversity of views fosters healthy progress and prevents stagnation. In September, I had the privilege of founding an interdisciplinary centre, the Black Hole Initiative at Harvard University in Cambridge, Massachusetts, which brings together astronomers, physicists, mathematicians and philosophers. Our experience is that a mix of scholars with different vocabularies and comfort zones can cultivate innovation and research outside the box. Already the centre has prompted exciting insights on the reality of naked singularities in space-time, the prospects for imaging black-hole silhouettes and the information paradox.

Such simple, off-the-shelf remedies could help us to avoid the scientific fate of the otherwise admirable Mayan civilization. ■

Avi Loeb is professor of science and chair of the astronomy department at Harvard University in Cambridge, Massachusetts, USA.

e-mail: aloeb@cfa.harvard.edu

1. Weinberg, S. *Phys. Rev. Lett.* **59**, 2607–2610 (1987).
2. Vilenkin, A. *Phys. Rev. Lett.* **74**, 846–849 (1995).
3. Bouso, R., Hall, L. J. & Nomura, Y. *Phys. Rev. D* **80**, 063510 (2009).
4. Loeb, A., Batista, R. A. & Sloan, D. J. *Cosmol. Astropart. Phys.* **8**, 040 (2016).
5. Hawking, S. W., Perry, M. J. & Strominger, A. *Phys. Rev. Lett.* **116**, 231301 (2016).
6. Ijjas, A., Steinhardt, P. J. & Loeb, A. Preprint at <https://arxiv.org/abs/1402.6980> (2014).
7. Dressing, C. D. & Charbonneau, D. *Astrophys. J.* **807**, 45 (2015).
8. Guillochon, J. & Loeb, A. *Astrophys. J. Lett.* **811**, L20 (2015).
9. Anglada-Escude, G. et al. *Nature* **536**, 437–440 (2016).



A lava flow from the Puu Oo volcanic cone in Hawaii.

Bridge the planetary divide

To explain why our planet is habitable, geoscientists studying Earth's surface and interior must work with each other and with communications scholars, write **Ariel D. Anbar, Christy B. Till and Mark A. Hannah.**

The classic 1970s British television series *Upstairs, Downstairs* is a good metaphor for our planet's evolution. Like the show, Earth's habitability depends on the dynamics of a complex household, and on subtle interactions between divided worlds.

Upstairs, at its surface, Earth is rich in molecular oxygen. O₂ is the second-most abundant gas in the atmosphere, making up 21% of our air. It reacts readily, so most

of Earth's surface is oxidized. Downstairs, by contrast, in Earth's interior, molecular O₂ is vanishingly rare. Materials brought up from the mantle, such as volcanic rocks, react with O₂ when they are exposed. Earth's oxidized surface is a veneer enveloping a vast O₂ sink.

This contrast was not always so stark. It changed halfway through the planet's history. Around 2.3 billion years ago, the amount of O₂ in the atmosphere rose ▶

► above one part per million, beginning an ascent to the high levels of today¹. This Great Oxidation Event transformed Earth and made intelligent life possible. Its cause remains a mystery. Solving it is a key challenge for Earth-systems scientists. It is also a challenge for astrobiologists: their ability to use O₂ as a signature of life on planets beyond our Solar System hinges on a better understanding of how it arose on Earth.

Key to that story is the balance between organisms' production of O₂ on the surface and consumption of the gas by reactions with rocks, fluids and gases from the interior. But we lack a quantitative theory of our planet's evolution that links changes at the surface with those below.

In part, that is because the surface and solid Earth research communities struggle to communicate with each other. After examining interactions within our own large Dynamics of Earth System Oxygenation team, funded by the US National Science Foundation, we learned that researchers in these neighbouring fields barely speak the same language. Our challenges are as much sociological as they are scientific.

A fuller theory of Earth evolution will emerge by bridging three divides. First, geoscientists studying the surface history of O₂ — typically geobiologists and low-temperature geochemists — need to understand how the gas is influenced by what goes on below. Second, geoscientists studying Earth's interior — geophysicists and high-temperature geochemists — must realize that such questions are also germane to some of their most vexing challenges. Third, geoscientists of all stripes should improve their conversations by integrating methods from communications disciplines.

SURFACE PUZZLES

Geoscientists trying to explain the rise of O₂ in Earth's atmosphere increasingly realize that something prevented this gas from accumulating for a billion years or more before the Great Oxidation Event. Geological evidence is mounting that photosynthesis was producing O₂ as early as 3.5 billion years ago. Microbial 'mats' in shallow waters at ancient seashores — preserved today as fossilized stromatolites that date back to at least that time — could have been inhabited by O₂-making cyanobacteria. Studies of the abundances of carbon, molybdenum and other elements and their isotopes in marine shales and carbonate rocks support this picture¹.

Interactions between the surface and

interior of the Earth are implicated². Rocks derived from the mantle, such as basalts, consume O₂ when they weather. Oxygen also reacts with hydrogen and other gases released from volcanoes, hydrothermal vents and mineral reactions. Because the atmosphere is thin compared with the planet's internal bulk, even small changes in the rates at which these rocks and gases consume O₂ could have a big impact. Those changes might arise from alterations in the compositions of materials derived from the mantle, or in the rates at which they are brought to the surface or are dragged back down by the subduction of tectonic plates.

COOLING TROUBLES

Many solid Earth scientists are unaware that the quest to understand the rise of O₂ in Earth's atmosphere can provide new impetus to investigations of fundamental aspects of the planet's internal evolution.

As Earth cooled after its formation, mantle convection may have slowed. The abundance of iron and magnesium in magmas derived from the mantle decreased. The modern tectonic processes that recycle crust into the mantle kicked in. And the crust became richer in silicon dioxide (SiO₂), and more buoyant^{3–5}. The distribution of heat and elements in the mantle were altered as surface minerals were mixed in and as iron–nickel alloy was steadily lost to Earth's growing core.

Such cascades of changes probably affected O₂ at the surface. By the time of the Great Oxidation Event, the rate of O₂ consumption by reaction with rocks and gases originating in Earth's interior may have slowed enough that O₂ produced by photosynthesis could accumulate in the atmosphere.

None of these changes is well-quantified. For instance, whether the upper mantle's capacity to consume O₂ evolved or not is still debated. For 20 years, researchers thought that it did not. But recent measurements (some conducted by members of our team) of vanadium and scandium in ancient rocks derived from the mantle indicate that its O₂ consumption capacity might have fallen in the 1.5-billion-year run-up to the Great Oxidation Event^{6,7}. Changes in the composition of the continental crust also suggest a decrease in O₂ consumption by rock-weathering processes around the same time⁴.

Thus, unravelling the mystery of Earth's O₂ requires a quantitative theoretical model of the physics and chemistry of planetary cooling and its consequences for surface–interior interactions over time.

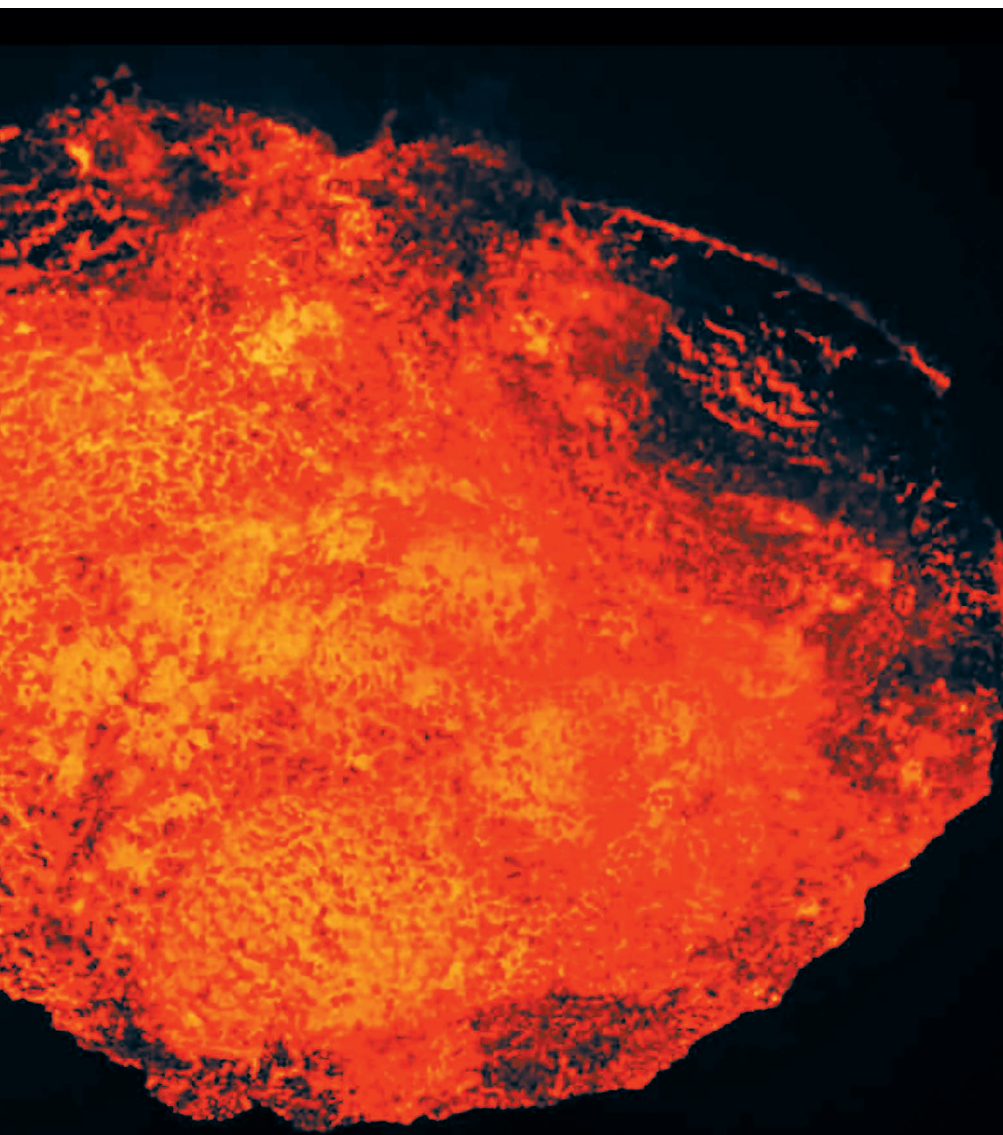
LOOKING BEYOND

Such a model would have benefits beyond the geosciences. Astronomers are hoping to use O₂ as a fingerprint of life on



Earth-like exoplanets. But will O₂ inevitably accumulate if biological processes produce it in large amounts? Stars have a wide range of abundances of elements such as carbon, magnesium and silicon. The exoplanets that form around them must vary in their compositions too, which would affect their tectonics and internal chemistry⁸. So, on some worlds, the rate of surface–interior interactions or the mantle's capacity to consume O₂ may be so high that the gas cannot accumulate. On such worlds, O₂ may be useless as a signature of life.

Astronomers need to know which exoplanets are worth investigating intensely for O₂, and for which this might be a waste of precious telescope time. A quantitative model that incorporates a wide range of planetary compositions would indicate which Earth-like planets have a chance of developing O₂-rich atmospheres and which will never do so even if they are teeming with O₂-producing life.



Explorer Sam Crossman descends into Vanuatu's fiery Marum lava lake.

is stalled by even a seemingly simple question such as: 'How can we compare a sediment's capacity to consume O_2 relative to the mantle?'

Scholars who study how people share ideas have analytical skills and methods that can address this challenge. These begin with carrying out surveys and interviews, and designing visualizations to demonstrate differences in use of language and its impacts (see Supplementary Information; go.nature.com/2e0gyip).

Such data feed into analyses of social networks that help team leaders to identify and empower investigators most able to bridge subdisciplines — in our case, people who score highly on understanding both surface and deep Earth terms — or to identify people with hybrid knowledge who can address particular points of overlap. Because investigators are attuned mainly to their own group's language, efforts must be made to help each group appreciate how their concepts relate to others' and how each perspective informs the research questions pursued.

By gaining such awareness and working together more effectively, geoscientists studying Earth's surface and interior, drawing on analyses of discourses and team dynamics, can build a model for the evolution of Earth's O_2 rise, better understand the history of Earth's habitability, and inform the search for life on worlds beyond our own. ■

Ariel D. Anbar is professor, and **Christy B. Till** is assistant professor, in the School of Earth & Space Exploration, Arizona State University, Tempe, Arizona, USA. **Mark A. Hannah** is assistant professor in the Department of English, Arizona State University, Tempe, Arizona, USA. e-mail: anbar@asu.edu

- Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. *Nature* **506**, 307–315 (2014).
- Catling, D. C. in *Treatise on Geochemistry* 2nd edn (eds Holland, H. D. & Turekian, K. K.) 177–195 (Elsevier, 2014).
- Condie, K. C. *Chem. Geol.* **104**, 1–37 (1993).
- Gaschnig, R. M. et al. *Geochim. Cosmochim. Acta* **186**, 316–343 (2016).
- Tang, M., Chen, K. & Rudnick, R. L. *Science* **351**, 372–375 (2016).
- Aulbach, S. & Stagno, V. *Geology* **44**, 751–754 (2016).
- Nicklas, R., Puchtel, I. & Ash, R. 'The oxidation state of Archean komatiites revisited' *Goldschmidt Abstracts* 2272 (2015).
- Young, P. A. et al. *Astrobiology* **14**, 603–626 (2014).
- Milligan, R. A., Gilroy, J., Katz, K. S., Rodan, M. F. & Subramanian, K. S. *Holistic Nursing Pract.* **13**, 47–53 (1999).
- Hannah, M. A. & Lam, C. *Technical Commun.* **63**, 328–345 (2016).

Supplementary information accompanies this article online at go.nature.com/2e0gyip.

CROSSING DIVIDES

To bridge these disciplinary divides takes effective conversation. Yet few scientists and engineers realize how deeply language affects their collaborations and research. To close this gap, our team developed a research partnership with some social scientists and humanities scholars who work on communication and team dynamics. Research in their field shows how diverse teams work more effectively when they develop a shared language — common vocabulary, jargon, codes and linguistic styles as well as implicit understandings^{9,10}.

Our first step was to examine the language that our investigators used, to identify and confront gaps in our group's understanding of concepts related to Earth's oxygenation. We were motivated by a paradox that we observed: scientists in closely allied disciplines find it hardest to communicate effectively with one another. Astronomers, biologists and geoscientists are willing to ask each other 'dumb'

questions that expose shared and divergent understanding. But a geobiologist working with a geophysicist might assume a shared understanding that does not exist.

For example, solid Earth scientists and geobiologists share the word 'oxygenation' but in fact lack a common language to

"Scientists in closely allied disciplines find it hardest to speak to one another."

describe the amount of O_2 that is available to react. Geobiologists, used to high atmospheric levels, think in terms of O_2 partial pressures and molarities when dissolved in solution, and so have developed a specialized vocabulary to describe environments with different amounts of free O_2 (such as 'oxic', 'anoxic', 'suboxic' and 'euxinic'). Solid Earth scientists use the physio-chemical term, 'oxygen fugacity', to reflect the fact that oxygen in the deep Earth is mainly locked in minerals and not in the form of an ideal gas. So conversation



Pages from the 'herbal' section of the Voynich manuscript.

CRYPTOGRAPHY

Calligraphic conundrum

Andrew Robinson relishes a new volume on a work that has long defied decoders.

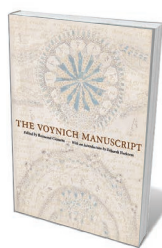
In a Connecticut archive sits a manuscript justifiably called the most mysterious in the world. Since its rediscovery more than a century ago, the Voynich manuscript has been puzzled over by experts ranging from leading US military cryptographer William Friedman to cautious (and incautious) humanities scholars. Since 1969, it has been stored in Yale University's Beinecke Rare Book and Manuscript Library in New Haven.

The fine calligraphy of the 234-page 'MS 408', apparently alphabetic, has never been decoded. Copious illustrations of bathing women, semi-recognizable plants and apparent star maps remain undeciphered. No one knows who created it or where, and there is no reliable history of ownership. Its parchment was radiocarbon-dated in 2009 to between 1404 and 1438, with 95% probability. The manuscript could still be a forgery using medieval parchment, but most experts, including Yale's, are convinced it is genuine.

Now *The Voynich Manuscript*, a volume edited by the library's curator, Raymond Clemens, revivifies this tantalizing artefact. A handsome facsimile is accompanied by an introduction by historian of science Deborah Harkness and six up-to-date essays by conservators, historians and literary scholars. As Harkness remarks, the collection does not

attempt a definitive solution of the conundrums raised by the manuscript. Instead, the contributors "invite the reader to join us at the heart of the mystery as we strive to better understand this complex book and its history".

The manuscript is named after Wilfrid Voynich, a Polish-born revolutionary who escaped Siberian exile to become a dealer in rare books and manuscripts in London and then New York City. In 1912, under condition of absolute secrecy, Voynich bought the manuscript from a Jesuit archive in Italy that he never identified, but that was selling part of its collection to the Vatican Library. Up to his death in 1930, Voynich marketed it as the work of thirteenth-century English scientist and friar Roger Bacon — a theory discredited only by the radiocarbon dating. In 1961, the manuscript was purchased by book dealer Hans Kraus, who presented it to Yale; the university put it online in 2004 (see go.nature.com/2dlns1c). Today, it attracts



The Voynich Manuscript
EDITED BY RAYMOND CLEMENS
Yale University Press: 2016.

16% of traffic to the Beinecke digital library, and for close to half of all traffic to the zoom viewer that allows examination of individual pages of manuscripts. When the late semiotician and novelist Umberto Eco visited Yale in 2013, MS 408 was the only manuscript he asked to see, notes Clemens.

The high-quality colour facsimile makes up most of the book. Each page, including foldouts, is reproduced at almost its original size (around 23 × 16 centimetres). Among the essays, researcher René Zandbergen covers the disputed history of the manuscript's ownership from the fifteenth century to 1912. Manuscript curator Arnold Hunt looks at Voynich's respected, but not always scrupulous, career. A group of six conservators details the forensic investigation of the parchment, ink and binding. And historian Jennifer Rampling probes the relationship of the illustrations to those in alchemical manuscripts, finding "no clear parallel" in alchemical writing to the predominance of female bathers.

The story of the various failed attempts to decipher the script, told by Clemens and Renaissance scholar William Sherman, is particularly fascinating. It begins in the 1920s, when US philosopher William Newbold convinced himself that the text was meaningless, but that each letter concealed an

BEINECKE RARE BOOK AND MANUSCRIPT LIB., YALE UNIV.

SSPL/GETTY

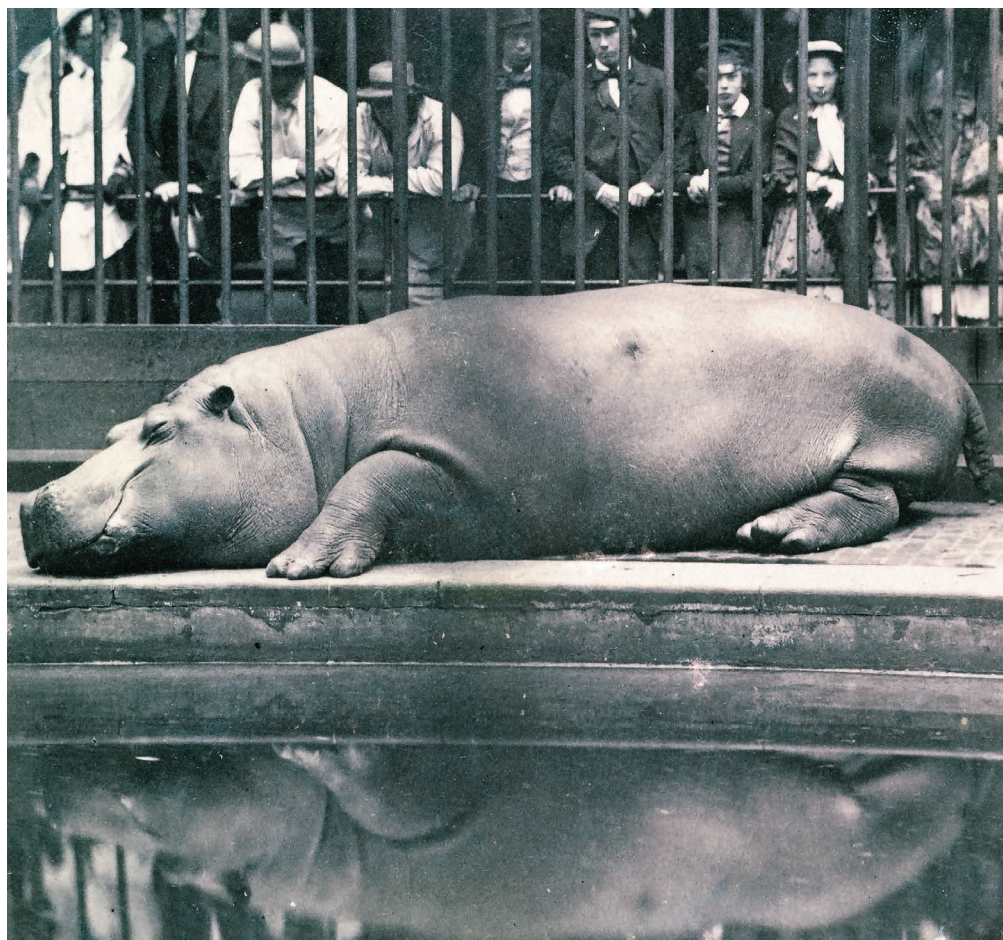
ancient Greek shorthand readable under magnification. He further claimed that this 'finding' proved the authorship of Bacon, who he claimed had invented a microscope centuries before Antonie van Leeuwenhoek. After Newbold's death, the 'shorthand' was revealed to be random cracks left by drying ink.

Wisely, little space is devoted to the many speculative theories of origin and meaning. The manuscript has been cast, for instance, as a Middle High German hygiene manual written in 'mirror writing' — the technique used by Leonardo da Vinci — and as a herbal manuscript in the Aztec language Nahuatl. (Readers with a taste for these can consult *The Voynich Manuscript* (Orion, 2004), a study by Gerry Kennedy and Rob Churchill.) And no new decipherment is offered.

Some idea of the complexity of the story is shown by a letter in Latin that Voynich apparently found affixed to the manuscript. This is reproduced in the volume, oddly without translation. Dated 1665, it was written by Johannes Marcus Marci (physician to the Holy Roman Emperors) and addressed to his former tutor, the Jesuit Athanasius Kircher. (The foremost polymath of the age, Kircher was wrongly credited with deciphering Egyptian hieroglyphs.) In the letter, Marci notes that he is sending Kircher the entire manuscript, asks him to decode it and mentions the claim that Bacon authored it. We know that Kircher received the manuscript, but made no progress with it. After his death in 1680, it disappeared into Jesuit archives in Rome until it came into Voynich's hands.

What hope is there of decoding the script? Not much at present, I fear. The Voynich manuscript reminds me of another uncracked script, on the Phaistos disc from Minoan Crete, discovered in 1908. The manuscript offers much more text to analyse than does the disc, but in each case there is only one sample to work with, and no reliable clue as to the underlying language — no equivalent of the Rosetta Stone (A. Robinson *Nature* **483**, 27–28; 2012). Professional cryptographers have been rightly wary of the Voynich manuscript ever since the disastrous self-delusion of Newbold. But inevitably, many sleuths will continue to attack the problem from various angles, aided by this excellent facsimile. Wide margins are deliberately provided for readers' notes on their own ideas. "Bonne chance!" writes Clemens. I'll second that. ■

Andrew Robinson's many books include *The Man Who Deciphered Linear B and Lost Languages*.
e-mail: andrew@andrew-robinson.org



Obaysch the hippo was captured in 1849 and sent to London Zoo, where he became a sensation.

ZOOLOGY

Animal crackers

Henry Nicholls relishes a brace of chronicles on how zoos on both sides of the Atlantic came to be.

Stamford Raffles did not waste his time. In 1825, little more than six months after returning to London from the East Indies, he'd put together a prospectus that would result in the creation of the Zoological Society of London (ZSL). With his career as an entrepreneur-cum-statesman in Penang, Java, Sumatra and Singapore behind him, Raffles was ready to indulge his passion for natural history.

The relationship between humans and the rest of the animal kingdom has always changed and will always change. But there can be few shifts as rapid and radical as those in the nineteenth century. With the age of sail in full swing and European docksides piled with boxes of specimens, a new class of professional zoologist arose. The likes of Alexander von Humboldt, Charles Darwin and Alfred Russel Wallace began to make sense of the

The Zoo: The Wild and Wonderful Tale of the Founding of London Zoo

ISOBEL CHARMAN
Viking: 2016.

The Animal Game: Searching for Wildness at the American Zoo

DANIEL E. BENDER
Harvard University Press: 2016.

astounding variety of animal life. The period covered in Isobel Charman's *The Zoo*, 1824–51, saw much of the transformative action. Meanwhile, historian Daniel Bender's *The Animal Game* chronicles the evolution of the US zoo from the 1870s to the 1970s.

Charman has hit on a delightful structure for her "wild and wonderful tale". Each chapter is a leg in a relay. So Raffles hands over to Decimus Burton, the ambitious twenty-something architect who began to shape

► the zoological gardens in Regent's Park in 1827. Veterinary surgeon Charles Spooner is next, his struggle to keep the animals alive in the 1830s foreshadowing concerns over animal welfare. On the upside, the deceased creatures were a bonus for chief animal preserver John Gould and the ZSL's museum in Mayfair, a collection subsumed into the British Museum in 1855. Gould pops up again later with a selection of Galapagos finches, which helped Darwin to develop the case for evolution by natural selection. Thus Charman takes the story out of the cages and onto the smoggy, sometimes riotous streets of Victorian London, up and down the country and beyond its shores.

With an imperial network of travellers and traders, customs officials and collectors, dealers and diplomats all sending specimens to London, the nascent zoo boasted an impressive diversity of animals from the off. During its first decades, the zoological gardens remained relatively exclusive, as its fellows mulled over the peculiar physiology, morphology and behaviour of inmates such as a chimpanzee that they named Tommy. The public, meanwhile, indulged their curiosity at venerable but unashamedly unscientific menageries such as the Exeter Exchange on London's the Strand. Soon, Raffles' original vision, emphasizing the animals as objects of scientific research, became hard to sustain. The ZSL gave up its farm outside Kingston upon Thames — a site for breeding and experiments — and opened the gates of the zoological gardens in the 1840s. It also began to bring in crowd-pleasers such as giraffes and hippopotamuses.

What is perhaps most striking about *The Zoo* is its style. "He could see it now, in his mind's eye," Charman writes as Raffles' plans take shape. "All of these beasts, from across the earth, from every corner of the endless Empire, gathered right here! He could almost hear the roars." It is clear that a huge amount of research has gone into this work, but much is lost among imagined thoughts and feelings. In writing non-fiction novelistically, Charman fails to take full advantage of the strengths of either genre.

Bender takes a more conservative approach in *The Animal Game*. Given the imperial premise of a robust zoological collection, it makes perfect sense that the appearance of US zoos should map onto the emergence of the nation as a global power towards the end of the nineteenth century. The Philadelphia Zoological Society — the country's oldest — appeared in 1859, although it achieved a permanent collection

"Urban elites who dreamed of zoological parks fantasized that displays of biological order would beget social order."



William Mann, fifth director of the National Zoo in Washington DC, travelled widely to collect animals.

(including an Asian elephant, Jennie) only some 15 years later. Others included Lincoln Park Zoo (Chicago, Illinois, in 1868), the National Zoo (Washington DC, 1889) and the Bronx Zoological Park in New York City (1899). In many ways, the stories of these institutions mirror the ZSL's. As with Raffles, "those urban elites who dreamed of zoological parks fantasized that displays of biological order would beget social order", writes Bender: nature's hierarchy offered a model to the masses. Also like the ZSL, US zoological societies struggled to distinguish themselves from low-brow entertainment — menageries, circuses and world fairs.

Whereas the animal trade is implicit in *The Zoo*, Bender's book renders it explicit through characters such as Frank Buck — animal dealer, showman and film star — and William Mann, the pith-helmeted director of the National Zoo from 1925 to 1956. In 1937, Mann travelled to the Far East to secure tigers (at US\$100 a pair), a Sumatran gibbon, cassowaries, orang-utans and more. In Africa, he had to manage his own collecting, on one occasion employing 500 locals in a failed attempt to net a giraffe. When he sailed for the United States, he had a cargo of 1,500 animals, but many were dying: a crate of 14 pythons was summarily thrown overboard.

➔ **NATURE.COM**
For more on science
in culture see:
[nature.com/
booksandarts](http://nature.com/booksandarts)

During the Great Depression, bankrolled by federal coffers, many zoos began to reconfigure with new displays such as

"monkey islands" surrounded by moats. Captive breeding eventually appeared, as much by necessity as design: by the 1960s, centuries of exploitation had devastated wild animal populations. The only way that zoos could survive was to put more effort into breeding their own animals.

Bender zips back and forth from institution to institution, on collecting trips with traders, and with visitors feeding the animals snacks (and on occasion, broken bottles and roofing nails). He draws on rich archival material, including the thoughts of key players such as zoologist William Temple Hornaday, memos from management and a flurry of clippings from the popular press. In the century that he covers, US zoos faced challenges from economic ups and downs to rogue animals, disgruntled staff and a demanding public.

What emerges is a story of adaptation and survival that exposes the modern zoo as "a third nature", a "product of how we built, lived, and contested empires. It is wildness and wilderness suspended at the moment of their initial enclosure when there were still plenty of animals for the taking." Those who are ethically opposed to zoos will find plenty here to strengthen their case. But with zoos' power of reinvention, it seems likely that this "third nature" will be with us for some time. ■

Henry Nicholls is a journalist based in London. His latest book is *The Galapagos*. e-mail: henry@henrynicholls.com

TONY LINCKY/THE LIFE PICTURE COLLECTION/GETTY

Correspondence

Norway wolf cull will hit genetic diversity

Norway's regional management authorities have approved plans to cull up to 70% of its grey wolves (*Canis lupus*). Because the current population consists of just 65–68 individuals, of which only 21 are thought to be sexually mature (see go.nature.com/2euyixy), this cull and the subsequent loss of genetic material would seem to be a significant misstep for ensuring the wolf's persistence in Norway.

Preserving genetic variation is crucial for the long-term survival of populations in fragmented landscapes. The genetic diversity of these wolves is already low because the entire population is descended from just a few individuals (C. Vilà *et al. Proc. Biol. Sci.* **270**, 91–97; 2003).

Their small population size and founding history mean that they need genetic input from immigrant wolves from Sweden, Finland and Russia to survive. Otherwise, they face the same fate as the well-studied Isle Royale wolf population in the United States (see *Nature* **520**, 415; 2015).

Livestock protection is the Norwegian government's main justification for the cull, although less than 9% of Norway's sheep are taken by wolves (data from go.nature.com/2eemukj). The government is ignoring well-established scientific practices for managing a critically threatened species, as well as overriding its commitment to the Bern Convention, which lists the wolf as a strictly protected species.

Elina Immonen *Uppsala University, Sweden.*

Arild Husby *University of Helsinki, Finland.*
arild.husby@helsinki.fi

Inequality: need for data on all nations

Inequality studies need to be more representative — the countries that suffer the most from inequality are also those that we know the least about (see

Nature **537**, 466–470; 2016).

Between 1994 and 2013, North America and western Europe together accounted for 80% of all publications on inequalities and social justice (ISSC, IDS and UNESCO *World Social Science Report*; UNESCO, 2016). The number of publications and data on these themes are very low from sub-Saharan Africa, south and east Asia, former eastern-bloc countries and the Arab world. This imbalance in our knowledge about inequality diminishes our global capacity to address it.

Mathieu Denis *International Social Science Council, Paris, France.*

Melissa Leach *Institute of Development Studies, University of Sussex, Brighton, UK.*
mathieu@worldsocialscience.org

Inequality: span the global divide

National initiatives need to correct injustices related to class, inequality and salaries among scientists (see *Nature* **537**, 466–470; 2016). However, such measures may serve to reinforce the global north–south divide in research if, perhaps inevitably, they are more prevalent in higher-income countries.

The domination of the scientific agenda and literature by northern over southern researchers has serious implications for how science is designed and produced, undermining its salience, credibility and legitimacy — and therefore its influence on policy development and implementation (see D. W. Cash *et al. Proc. Natl Acad. Sci. USA* **100**, 8086–8091; 2003).

We suggest that the world's many transboundary issues — such as climate change, poverty, human migration, public health and biodiversity decline — call for a more comprehensive, global approach. This should span the north–south divide by addressing the underlying issues and their consequences (see, for example, M. Blicharska *et al.*

Nature Clim. Change; in the press).

Richard J. Smithers *Ricardo Energy & Environment, Harwell, UK.*

Malgorzata Blicharska *Uppsala University, Uppsala, Sweden.*

José María Gutiérrez *University of Costa Rica, San José, Costa Rica.*
richard.smithers@ricardo.com

Protect aquaculture from ship pathogens

Aquaculture is the world's fastest-growing food-production sector and a crucial contributor to the United Nations' Sustainable Development Goals. As a group of scientists, ocean-policy experts, aquaculture professionals and technical consultants from international organizations, we argue that, despite recent legislation, fish farms may still be at risk from pathogens in ballast water discharged from ships.

The International Convention for the Control and Management of Ships' Ballast Water and Sediments (go.nature.com/2evuskh) will come into force in September 2017. This will reduce the risks of transfer of organisms larger than 10 micrometres and of bacteria that are harmful to humans, including *Vibrio cholerae*, *Escherichia coli* and *Enterococcus* species. However, the convention does not mention any other aquatic bacteria or viruses that could cause epidemics in the US\$160-billion aquaculture industry and threaten food security.

We suggest using a combination of molecular tools, experimental investigation, monitoring data and operational models to evaluate the risks and possible impacts on local aquaculture farms of how, when and where ships discharge ballast water. The findings should be presented to UN-OCEANS (www.unoceans.org) and the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (www.gesamp.org) to support

existing regulatory frameworks.

Guillaume Drillet* *DHI, Singapore.*

gdr@dhigroup.com

*On behalf of 11 correspondents (see go.nature.com/2en2btv for full list).

Small data call for big ideas

Despite the meteoric rise in big-data approaches, funders also need to recognize that some of the most pressing problems must instead rely on the intelligent use of small data sets.

Value-of-information analysis can evaluate whether big-data collection is worthwhile (R. Schlaifer and H. Raiffa *Applied Statistical Decision Theory*; Clinton, 1961). Collecting big monitoring data for threatened or invasive species, for instance, risks delaying decisions on protective measures. It might be better to fund direct action, as for killer whales in the Georgia basin (E. McDonald-Madden *et al. Trends Ecol. Evol.* **25**, 547–550; 2010).

Urgent decisions may be necessary when information is sparse. In agricultural systems, a fast response to a new pest or disease outbreak can make the difference between eradication and decades of costly quarantine programmes. In ecology, population sizes and detectability are often too low to create big data sets. In health, defence and social sciences, collecting big data risks violating human ethics.

Where data are limited, scientific solutions underpinned by strategies such as adaptive management can optimize decision making (I. Chadès *et al. Theor. Ecol.* <http://doi.org/br9s>; 2016). Artificial intelligence, for example, can inform adaptation strategies for sea-level rise to protect migratory shorebirds worldwide (S. Nicol *et al. Proc. R. Soc. B* **282**, 20142984; 2015).

Iadine Chadès, Sam Nicol *CSIRO, Dutton Park, Queensland, Australia.*
iadine.chades@csiro.au

The intrinsic thermal conductivity of SnSe

ARISING FROM L.-D. Zhao *et al.* *Nature* **508**, 373–377 (2014); doi:10.1038/nature13184

Several groups have been unable to reproduce the record high thermoelectric figure of merit ZT of SnSe reported in ref. 1. Zhao *et al.*¹ measured an ultralow thermal conductivity ($<0.4\text{ W m}^{-1}\text{ K}^{-1}$ at 923 K) and consequently record high values of $ZT \approx 2.6 \pm 0.3$ and $ZT \approx 2.3 \pm 0.3$ at 923 K along the b and c directions, respectively, in their single-crystalline SnSe. However, after careful analysis of the data of ref. 1, we deduce that their samples are not fully dense and thus not truly single crystalline, implying that their reported thermal conductivities are not intrinsic to SnSe. This warrants further investigation into intrinsic thermal transport in SnSe single crystals and its use as a thermoelectric material. There is a Reply to this Comment by Zhao, L.-D. *et al.* *Nature* **539**, <http://dx.doi.org/10.1038/nature19833> (2016).

In ref. 1, the total thermal conductivity κ of single-crystalline SnSe was calculated using the relation $\kappa = DC_p\rho$, where D is the thermal diffusivity, C_p is the specific heat capacity at constant pressure, and ρ is the density. Although the authors did not list the density of their SnSe crystals, they did provide the diffusivity, specific heat and total thermal conductivity data along the three major crystallographic directions (source data for figure 2d, and extended data figure 6a and b, of ref. 1). We extracted the ρ value for their crystals using the above relation, and found their single-crystalline SnSe samples to be of much lower density (around 5.43 g cm^{-3}) when compared to the theoretical density ρ_{th} reported in the literature (Table 1). In their Reply to this Brief Communication Arising, Zhao *et al.* confirmed that the experimentally measured density of their samples was indeed lower than ρ_{th} . The ρ value of single-crystalline SnSe estimated from the powder X-ray diffraction studies were in the range $6.13\text{--}6.18\text{ g cm}^{-3}$ (Joint Committee on Powder Diffraction Standards (JCPDS) card numbers 01-089-232, 01-089-233 and 01-089-235)², while neutron diffraction studies³ reported a ρ value of about 6.18 g cm^{-3} (JCPDS card number 01-071-3877). In addition, electron diffraction studies reported a ρ value of 6.07 g cm^{-3} (JCPDS card number 01-075-2123)⁴. In other words, the ρ value for the SnSe samples used in ref. 1 was 88% to 90% of the theoretical density ρ_{th} .

Although the exact cause of the low ρ value in their SnSe samples is beyond our conjecture, in Supplementary Table 1 we compare the reported density and the thermal conductivity values given by several groups for SnSe at 300 K and 750 K, measured along different pressing directions or crystallographic axes. As can be deduced from Supplementary Table 1 and the corresponding scatter plot (Fig. 1), the κ values and the corresponding densities of fully dense single and polycrystalline SnSe reported by different groups are consistently higher than those reported in ref. 1. In the case of polycrystalline SnSe, additional phonon scattering mechanisms are present, which should lead to even lower lattice thermal conductivity κ_{L} ($\approx \kappa$) compared to the corresponding thermal conductivity of single-crystalline SnSe. Although Zhao *et al.*, in their Reply to this

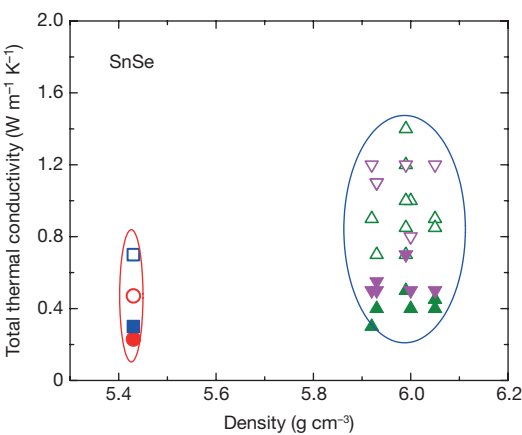


Figure 1 | A scatter plot of thermal conductivity versus density of single and polycrystalline SnSe. The values are taken along different crystallographic axes and along parallel or perpendicular pressing directions, at 300 K and about 760 K, derived from the peer-reviewed literature cited in Supplementary Table 1. The open and solid symbols represent κ at 300 K and 760 K, respectively. The κ values of single-crystalline SnSe along the a axis are represented by red circles and along the b and c axes are represented by blue squares. The κ values of polycrystalline SnSe along the parallel and perpendicular pressing directions are represented by green triangles and violet inverted triangles, respectively. The red ellipse encompasses data from ref. 1 and data reported by other groups^{6–21} lie within the blue ellipse. A detailed list of all the data points is available in Supplementary Table 1.

Brief Communication Arising, suggested the formation of SnO_2 as a plausible cause for the higher thermal conductivity in polycrystalline SnSe, it must be noted that SnSe is not highly sensitive to air at ambient temperature⁵.

It should also be noted that Zhao *et al.*¹ suggested the presence of “plenty of vacancies and interstitials” in their samples. Some defects are entropically present in single crystals, but we doubt that the vacancies and interstitials can account for the observed 11%–12% deficiency in density at room temperature. It is well known that porosity significantly reduces the thermal transport, which emphasizes the importance of reporting packing density values in future publications to validate the intrinsic transport properties. In other words, true thermal conductivity cannot be obtained by simple normalization, because thermal diffusivity and density are interdependent quantities. Here, we are not expressing concerns about the measurement techniques or the self-consistency of measurements by Zhao *et al.*¹ but about the intrinsic nature of their SnSe single crystals.

Thermoelectricians have long aimed to optimize ZT by reducing κ , and ref. 1 reported exceptionally low thermal conductivity. Thus, the sole aim of this Comment is to correct the scientific record by stating that the ultralow κ value reported in ref. 1 is not intrinsic to fully dense single-crystalline SnSe. A single crystal, by definition, must have an experimentally measured density that is close to 100% of the theoretical density. Thus, the SnSe samples of ref. 1 cannot be classified as single crystalline and the thermal conductivity and figure of merit values for SnSe presented in ref. 1 are not intrinsic to single-crystalline SnSe.

Table 1 | Room-temperature characteristics of the SnSe samples of ref. 1

At about 300 K	a axis	b axis	c axis
Thermal conductivity, κ ($\text{W m}^{-1}\text{ K}^{-1}$)	0.46525	0.70014	0.67560
Diffusivity, D ($\text{mm}^2\text{ s}^{-1}$)	0.33983	0.51139	0.49347
Specific heat, C_p ($\text{J g}^{-1}\text{ K}^{-1}$)	0.252	0.252	0.252
Density, ρ (g cm^{-3})	5.43	5.43	5.43

The densities of the samples of ref. 1 were deduced from the relation $\kappa = DC_p\rho$.

Pai-Chun Wei¹, S. Bhattacharya², J. He², S. Neeleshwar³, R. Podila², Y. Y. Chen¹ & A. M. Rao²

¹Institute of Physics, Academia Sinica, Taipei 11529, Taiwan.

email: chen2@phys.sinica.edu.tw

²Clemson Nanomaterials Institute and Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29634, USA.

email: arao@clemson.edu

email: bbhatta@g.clemson.edu

³University School of Basic and Applied Sciences, Guru Gobind Singh (GGS) Indraprastha University, New Delhi, India.

Received 17 January; accepted 31 August 2016.

1. Zhao, L. D. *et al.* Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals. *Nature* **508**, 373–377 (2014).
2. Adouby, K., Perez-Vicente, C. & Jumas, J. C. Structure and temperature transformation of SnSe. Stabilization of a new cubic phase Sn₄Bi₂Se₇. *Z. Kristallogr.* **213**, 343–349 (1998).
3. Chattopadhyay, T., Pannetier, J. & Von Schnering, H. G. Neutron diffraction study of the structural phase transition in SnS and SnSe. *J. Phys. Chem. Solids* **47**, 879–885 (1986).
4. Avilov, A. S., Imamov, R. M. & Navasardyan, S. N. Structure of thin films of SnS and SnSe. *Kristallografiya* **24**, 874–875 (1979).
5. Li, Y., He, B., Heremans, J. P. & Zhao, J.-C. High temperature oxidation behavior of thermoelectric SnSe. *J. Alloys Compd.* **669**, 224–231 (2016).
6. Wasscher, J. D., Albers, W. & Haas, C. Simple evaluation of the maximum thermoelectric figure of merit, with application to mixed crystals Sn_{1-x}Se_x. *Solid-State Electron.* **6**, 261–264 (1963).
7. Spitzer, D. P. Lattice thermal conductivity of semiconductors: a chemical bond approach. *J. Phys. Chem. Solids* **31**, 19–40 (1970).
8. Peng, K. *et al.* Broad temperature plateau for high ZTs in heavily doped p-type SnSe single crystals. *Energy Environ. Sci.* **9**, 454–460 (2016).
9. Zhao, L.-D. *et al.* Ultrahigh power factor and thermoelectric performance in hole-doped single crystal SnSe. *Science* **351**, 141–144 (2016).
10. Chen, C. L., Wang, H., Chen, Y. Y., Day, T. & Snyder, G. J. Thermoelectric properties of p-type polycrystalline SnSe doped with Ag. *J. Mater. Chem. A* **2**, 11171–11176 (2014).

11. Sassi, S. *et al.* Assessment of the thermoelectric performance of polycrystalline p-type SnSe. *Appl. Phys. Lett.* **104**, 212105 (2014).
12. Zhang, Q. *et al.* Studies on thermoelectric properties of n-type polycrystalline SnSe_{1-x}S_x by iodine doping. *Adv. Energy Mater.* **5**, 1500360 (2015).
13. Wei, T.-R. *et al.* Thermoelectric transport properties of pristine and Na-doped SnSe_{1-x}Te_x polycrystals. *Phys. Chem. Chem. Phys.* **17**, 30102–30109 (2015).
14. Li, Y., Shi, X., Ren, D., Chen, J. & Chen, L. Investigation of the anisotropic thermoelectric properties of oriented polycrystalline SnSe. *Energies* **8**, 6275–6285 (2015).
15. Leng, H.-Q., Zhou, M., Zhao, J., Han, Y.-M. & Li, L.-F. The thermoelectric performance of anisotropic SnSe doped with Na. *RSC Adv.* **6**, 9112–9116 (2016).
16. Leng, H., Zhou, M., Zhao, J., Han, Y. & Li, L. Optimization of thermoelectric performance of anisotropic Ag_xSn_{1-x}Se compounds. *J. Electron. Mater.* **45**, 527–534 (2016).
17. Han, Y.-M. *et al.* Thermoelectric performance of SnS and SnS-SnSe solid solution. *J. Mater. Chem. A* **3**, 4555–4559 (2015).
18. Serrano-Sánchez, F. *et al.* Record Seebeck coefficient and extremely low thermal conductivity in nanostructured SnSe. *Appl. Phys. Lett.* **106**, 083902 (2015).
19. Bera, T., Sanchela, A. V., Tomy, C. V. & Thakur, A. D. n-type SnSe_{1-x} for thermoelectric application. Preprint at <https://arxiv.org/abs/1601.00753> (2016).
20. Chere, E. K. *et al.* Studies on thermoelectric figure of merit of Na-doped p-type polycrystalline SnSe. *J. Mater. Chem. A* **4**, 1848–1854 (2016).
21. Ortiz, B. R. *et al.* Effect of extended strain fields on point defect phonon scattering in thermoelectric materials. *Phys. Chem. Chem. Phys.* **17**, 19410–19423 (2015).

Supplementary Information is available in the online version of the paper.

Author Contributions P.-C.W. and Y.Y.C. contributed towards the synthesis of fully dense SnSe single crystals and the thermoelectric measurements, which were verified by S.B., J.H., R.P., A.M.R. and S.N. This combined study helped us to determine that packing density has a substantial effect on the figure of merit of SnSe.

Competing Financial Interests Declared none.

doi:10.1038/nature19832

Zhao *et al.* reply

REPLYING TO P.-C. Wei *et al.* *Nature* **539**, <http://dx.doi.org/10.1038/nature19832> (2016)

In the accompanying Comment, Wei *et al.*¹ point out that the sample density used to obtain the thermal conductivity value of SnSe crystals is about 10% lower than the theoretical value ρ_{th} and as a result the thermal conductivity is underestimated by about 10%. The data published in ref. 2 are not based on a single measurement or on a single specimen. In the supplementary information of ref. 2 we reported measurements for at least seven crystals with good reproducibility. The material is not a conventional crystal and has several unusual and confusing features, such as the crystallographic phase transition (causing numerous microcracks) and a tendency to oxidation^{3–6}. The SnSe crystal also has strong anisotropy and weak mechanical properties, leading to facile cleavage so that the crystal can easily be damaged during cutting and handling. In addition, challenging measurements of this type (that is, thermoelectric measurements on cut single crystals) can show a statistical spread and the error in the measurement of thermal conductivity is typically between 15% and 20% (ref. 7).

The sample density we reported is the value that we measured. Differences between our value and the value(s) measured by others could be caused by differences in sample characteristics arising from preparation methods and handling, or measurement errors. Regardless of whether the density of the SnSe crystal is 6.0 g cm⁻³ or 5.5 g cm⁻³, the lattice thermal conductivity is ultralow, in the ranges 0.6–0.8 W mK⁻¹

at room temperature and 0.3–0.4 W mK⁻¹ at 800 K. In hole-doped SnSe crystals^{7,8}, where a different experimental density was measured (closer to the theoretical 6.0 g cm⁻³), the lattice thermal conductivity is in agreement with the value we report in ref. 2. If the experimental density is adjusted 10% upwards to be closer to the theoretical value of 6.1 g cm⁻³, the thermal conductivity is still ultralow. Therefore, sample density does not greatly affect the ultralow value of the lattice thermal conductivity. Whether the thermal conductivity is intrinsic to SnSe is not yet clear, but some explanations have recently been proposed^{3,5}.

The sample prepared by Wei *et al.* and described in their Comment was grown under different conditions, the details of which are not known to us. Must an ‘intrinsic’ value refer to a material with exactly stoichiometric composition, vacancy-free, dislocation-free, strain-free, fully dense, and with the number of excited electrons equal to the number of holes? Or could an ‘intrinsic’ value refer to the most stable state of a material for a given set of experimental growth conditions? The SnSe sample of a recent publication has an Sn vacancy rate of several per cent⁹, which is consistent with our observations of a large off-stoichiometry of Sn:Se (0.835:1) in SnSe crystals prepared using the Bridgman method; however, no off-stoichiometry was observed in SnSe crystals synthesized using a vapour-transfer method. Thus we consider it unfair to conclude, as do Wei *et al.* in their Comment, that our value for lattice thermal conductivity is not ‘intrinsic’ to SnSe, on the basis of the sample density only.

The true measure of the thermal conductivity lies in the thermal diffusivity data, which show very low values. The question of

interest concerning SnSe is whether or not it has an ultralow thermal conductivity, and most reported values for samples of SnSe suggest that it does^{2,3,10–14}. The reported values have a spread⁷, but this is not surprising given (1) the challenges in measuring materials with ultralow thermal conductivity, (2) variations from sample to sample with respect to defects, phase-transition cracks, and so on, and (3) several variations in the synthesis of the material⁵. Of all reports on polycrystalline SnSe samples^{2,3,7,8,10–36} at room temperature, 55% report higher thermal conductivity than in the single crystal², 25% report similar values, and 20% report a lower thermal conductivity. At 800 K, about 15% report higher thermal conductivity than in the single crystal², 80% report similar values, and 5% report a lower thermal conductivity.

Li-Dong Zhao¹, Shih-Han Lo², Yongsheng Zhang², Hui Sun³, Gangjian Tan¹, Ctirad Uher⁴, C. Wolverton², Vinayak P. Dravid² & Mercuri G. Kanatzidis^{1,2}

¹Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA.

email: m-kanatzidis@northwestern.edu

²Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA.

³Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA.

⁴Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA.

- Wei, P.-C. *et al.* The intrinsic thermal conductivity of SnSe. *Nature* **539**, <http://dx.doi.org/10.1038/nature19832> (2016).
- Zhao, L.-D. *et al.* Ultralow thermal conductivity and high thermoelectric figure of merit in SnSe crystals. *Nature* **508**, 373–377 (2014).
- Chen, Y.-X. *et al.* Understanding of the extremely low thermal conductivity in high performance polycrystalline SnSe through potassium doping. *Adv. Funct. Mater.* **26**, 6836–6845 (2016).
- de Kergommeaux, A. *et al.* Surface oxidation of tin chalcogenide nanocrystals revealed by ¹¹⁹Sn-Mössbauer spectroscopy. *J. Am. Chem. Soc.* **134**, 11659–11666 (2012).
- Zhao, L.-D., Chang, C., Tan, G. J., Kanatzidis, M. G. SnSe: a remarkable new thermoelectric material. *Energy Environ. Sci.* **9**, 3044–3060 (2016).
- Li, Y., He, B., Heremans, J. P. & Zhao, J. C. High-temperature oxidation behavior of thermoelectric SnSe. *J. Alloys Compd.* **669**, 224–231 (2016).
- Zhao, L.-D. *et al.* Ultrahigh power factor and thermoelectric performance in hole-doped single-crystal SnSe. *Science* **351**, 141–144 (2016).
- Peng, K. L. *et al.* Broad temperature plateau for high ZTs in heavily doped p-type SnSe single crystals. *Energy Environ. Sci.* **9**, 454–460 (2016).
- Sist, M., Zhang, J. & Brummerstedt Iversen, B. Crystal structure and phase transition of thermoelectric SnSe. *Acta Crystallogr. B* **72**, 310–316 (2016).
- Gharsallah, M. *et al.* Giant Seebeck effect in Ge-doped SnSe. *Sci. Rep.* **6**, 26774 (2016).
- Serrano-Sánchez, F. *et al.* Record Seebeck coefficient and extremely low thermal conductivity in nanostructured SnSe. *Appl. Phys. Lett.* **106**, 083902 (2015).
- Bera, T., Sanchela, A. V., Tomy, C. V. & Thakur, A. D. *n*-type SnSe_{1-x} for thermoelectric application. Preprint at <https://arxiv.org/abs/1601.00753> (2016).
- Kim, J. H. *et al.* Indium substitution effect on thermoelectric and optical properties of Sn_{1-x}In_xSe compounds. *J. Alloys Compd.* **682**, 785–790 (2016).
- Ju, H. & Kim, J. Effect of SiC ceramics on thermoelectric properties of SiC/SnSe composites for solid-state thermoelectric applications. *Ceram. Int.* **42**, 9550–9556 (2016).
- Kucek, V. *et al.* Thermoelectric properties of Ti-doped SnSe: a hint of phononic structure. *J. Electron. Mater.* **45**, 2943–2949 (2016).
- Singh, N. K. *et al.* The effect of doping on thermoelectric performance of p-type SnSe: promising thermoelectric material. *J. Alloys Compd.* **668**, 152–158 (2016).
- Wang, X. *et al.* Optimization of thermoelectric properties in *n*-type SnSe doped with BiCl₃. *Appl. Phys. Lett.* **108**, 083902 (2016).
- Chere, E. K. *et al.* Studies on thermoelectric figure of merit of Na-doped p-type polycrystalline SnSe. *J. Mater. Chem. A* **4**, 1848–1854 (2016).
- Fu, Y. J. *et al.* Enhanced thermoelectric performance in p-type polycrystalline SnSe benefiting from texture modulation. *J. Mater. Chem. C* **4**, 1201–1207 (2016).
- Leng, H. Q., Zhou, M., Zhao, J., Han, Y. M. & Li, L. F. The thermoelectric performance of anisotropic SnSe doped with Na. *RSC Adv.* **6**, 9112–9116 (2016).
- Leng, H. Q., Zhou, M., Zhao, J., Han, Y. M. & Li, L. F. Optimization of thermoelectric performance of anisotropic Ag_{0.1-x}Sn_{0.9-x}Se compounds. *J. Electron. Mater.* **45**, 527–534 (2016).
- Li, Y. W. *et al.* Enhanced mid-temperature thermoelectric performance of textured SnSe polycrystals made of solvothermally synthesized powders. *J. Mater. Chem. C* **4**, 2047–2055 (2016).
- Popuri, S. R. *et al.* Large thermoelectric power factors and impact of texturing on the thermal conductivity in polycrystalline SnSe. *J. Mater. Chem. C* **4**, 1685–1691 (2016).
- Tyagi, K. *et al.* Electrical transport and mechanical properties of thermoelectric tin selenide. *RSC Adv.* **6**, 11562–11569 (2016).
- Guan, X. H. *et al.* Thermoelectric properties of SnSe compound. *J. Alloys Compd.* **643**, 116–120 (2015).
- Ge, Z. H., Wei, K. Y., Lewis, H., Martin, J. & Nolas, G. S. Bottom-up processing and low temperature transport properties of polycrystalline SnSe. *J. Solid State Chem.* **225**, 354–358 (2015).
- Li, Y. L., Shi, X., Ren, D. D., Chen, J. K. & Chen, L. D. Investigation of the anisotropic thermoelectric properties of oriented polycrystalline SnSe. *Energies* **8**, 6275–6285 (2015).
- Zhang, Q. *et al.* Studies on thermoelectric properties of *n*-type polycrystalline SnSe_{1-x}S_x by iodine doping. *Adv. Energy Mater.* **5**, 1500360 (2015).
- Sassi, S. *et al.* Transport properties of polycrystalline p-type SnSe. *Mater. Today Proc.* **2**, 690–698 (2015).
- Wei, T. R. *et al.* Thermoelectric transport properties of pristine and Na-doped SnSe_{1-x}Te_x polycrystals. *Phys. Chem. Chem. Phys.* **17**, 30102–30109 (2015).
- Han, Y. M. *et al.* Thermoelectric performance of SnS and SnS-SnSe solid solution. *J. Mater. Chem. A* **3**, 4555–4559 (2015).
- Chen, C. L., Wang, H., Chen, Y. Y., Day, T. & Snyder, G. J. Thermoelectric properties of p-type polycrystalline SnSe doped with Ag. *J. Mater. Chem. A* **2**, 11171–11176 (2014).
- Sassi, S. *et al.* Assessment of the thermoelectric performance of polycrystalline p-type SnSe. *Appl. Phys. Lett.* **104**, 212105 (2014).
- Wei, T.-R. *et al.* Distinct impact of alkali-ion doping on electrical transport properties of thermoelectric p-type polycrystalline SnSe. *J. Am. Chem. Soc.* **138**, 8875–8882 (2016).
- Guo, H. *et al.* Enhanced thermoelectric performance of highly oriented polycrystalline SnSe based composites incorporated with SnTe nanoinclusions. *J. Alloys Compd.* **689**, 87–93 (2016).
- Morales Ferreiro, J. *et al.* Effect of the annealing on the power factor of un-doped cold-pressed SnSe. *Appl. Therm. Eng.* <http://dx.doi.org/10.1016/j.applthermaleng.2016.07.198> (in the press).

doi:10.1038/nature19833

BEHAVIOURAL BIOLOGY

Stones that could cause ripples

Monkeys have been observed pounding stones and unintentionally forming sharp-edged, tool-like fragments. This deliberate breakage raises questions about the evolution of intentional stone modification. [SEE LETTER P.85](#)

HÉLÈNE ROCHE

Of the non-human primates whose manipulative skills and tool use have been studied, capuchin monkeys (*Sapajus libidinosus*) seem to be the most. Although it has been known for more than 25 years that capuchins use tools¹, a key development in our understanding of their skills was the discovery² that, like chimpanzees, capuchins use stone tools for tasks such as cracking open nuts. On page 85, Proffitt *et al.*³ report an interesting stone-breaking behaviour observed among wild capuchins of Serra da Capivara National Park in eastern Brazil.

The authors observed capuchins deliberately damaging stones by using a hand-held quartzite cobble stone to violently hammer other quartzite cobbles embedded in a large stone conglomerate structure (Fig. 1). The stone-smashing activity (captured on video by Proffitt and colleagues) causes several types of stone modification. For example, the action can leave specific impact marks on the two cobbles, or the hand-held hammer-stone can fracture, creating chunky pieces or thinner detachments of sharp-edged stone flakes. The small concentrations of modified, broken and flaked stone pieces (known as a lithic assemblage) created by the capuchins' stone-breaking activity could potentially be mistaken for similar lithic assemblages made by early hominins.

Proffitt *et al.* describe the capuchin-produced lithic assemblages as being created in an unintentional way, because they are a coincidental consequence of the rock smashing. The authors report that the capuchins have never been seen using, or even seemed to be interested in, the stone flakes they produce (see Supplementary Video 1 in Proffitt *et al.*). However, Proffitt *et al.* consider that their capuchin findings are relevant to early-human studies because the process of stone knapping (striking stones together to make a desired stone product) is no longer a hominin-specific activity. Previous evolutionary explanations for the origins of hominin intentional stone modifications have focused on hominin-specific advances, such as changes in hand shape, coordination and cognitive skills. Explaining the origins of intentional stone modifications



T. FALÓTCÓ

Figure 1 | Capuchin monkeys can deliberately break stones. Proffitt *et al.*³ observed wild capuchin monkeys smashing stones together in Serra da Capivara National Park in Brazil. This process creates stone fragments, including some that have sharp edges and could potentially be used as tools — although the monkeys have not been observed to show interest in using these fragments.

by early hominins in light of Proffitt and colleagues' interpretation of their results would therefore require alternative hypotheses.

To understand the purpose of a given stone-modification activity, a key factor to consider is intention, a concept in the domain of cognitive psychology that implies having a goal to reach, involving planning and forethought. Similar stone-breakage outcomes can occur through activities that are either unintentional for the capuchins or intentional in the case of early hominins.

The stone-breaking action itself is a physical process subject to the permanent constraints of solid-state physics. Hard-rock conchoidal fracture, which produces flakes rather than shapeless chunks, is governed by the

mechanical laws of brittle-material fracture. Conchoidal fracture occurs when two stones are struck together at a tangential angle using a blow strength adapted to the stone density. Such fractures can occur naturally by chance. Animals can smash rocks together in the complex action required to produce conchoidally fractured fragments, regardless of whether or not the stone-smashing activity was carried out with the intention of stone modification.

The ability of New World capuchins to unintentionally produce lithic assemblages should not make us suspicious about any archaeological stone assemblage considered to belong to the African Early Stone Age. Our knowledge about technical behaviours and stone knapping in early-human archaeological

sites has a solid foundation. It represents decades of continuous research in the African continent that has provided hundreds of lithic assemblages, dated from as early as 3.3 million years ago⁴ and retrieved from sites that have a firmly established geological context. The African Early Stone Age archaeological record is based on a large body of evidence, including other contextual site data such as time-dating information and fossil analysis to consider together with the lithic assemblages. However, primate rock modification might bring new insight to the long-running archaeological debate about the origins and nature of modified stones identified from the American continent during the Late Pleistocene epoch (between 40,000 and 20,000 years ago)^{5–9}.

What early hominins and capuchins are doing to the stones is a goal-oriented action, but not with the same intention. In Early Stone Age archaeology, even if the function of all hominin stone tools is not known, stone-striking activities are mainly thought to be linked to intentional recurrent production of sharp-edged flakes. Other pounding activities producing less-distinctive stone artefacts might have occurred, but are less easy to demonstrate. Some stone tools from archaeological Early Stone Age sites have physical indications

of functional tool usage. Stone tools are often associated with faunal remains, such as bones bearing defleshing cut-marks inflicted by a sharp-edged tool, or the tools can have impact marks indicating breakage caused by a pounding tool¹⁰.

The goal of the capuchins' stone-pounding described by Proffitt and colleagues is unknown. The only obvious linked action is the monkeys' systematic licking of the pounded quartzite stone, which occurs after almost every blow to the pounded stone surface. Discussing the possibility of ingestion of nutritional components in the powdered rocks, the authors mention the trace nutrient silicon. However, silicon's essential role during connective-tissue formation and calcium deposition is at an early stage of bone calcification¹¹, yet one of the capuchins videoed in action by Proffitt and colleagues seems to be a young adult. Nevertheless, the benefit of ingesting powdered quartz is a possibility to consider.

The same research group recently demonstrated¹² that the capuchin monkeys of Serra da Capivara National Park have used quartzite cobbles to pound open nuts for at least 600 or 700 years. The capuchins might also have been engaged in unintentional knapping

for an equivalent or even longer period of time. Investigating the antiquity of the stone-smashing behaviour or trying to determine the behaviour's function and possible role in capuchin evolution are some of the many promising fields of research rippling out from the shattering discovery by Proffitt and colleagues. ■

Hélène Roche is in the UMR 7055 (PRETECH) of the CNRS, Paris-Nanterre University, 92023 Nanterre, France.
e-mail: helene.roche@mae.u-paris10.fr

1. Visalberghi, E. *Folia Primatol.* **54**, 146–154 (1990).
2. Fraguaszy, D. M., Izar, P., Visalberghi, E., Ottoni, E. B. & de Oliveira, M. G. *Am. J. Primatol.* **64**, 359–366 (2004).
3. Proffitt, T. *et al. Nature* **539**, 85–88 (2016).
4. Harmand, S. *et al. Nature* **521**, 310–315 (2015).
5. Lahaye, C. *et al. J. Archaeol. Sci.* **40**, 2840–2847 (2013).
6. Boëda, E. *et al. Antiquity* **88**, 927–941 (2014).
7. Aimola, G., Andrade, C., Mota, L. & Parenti, F. *J. Lith. Stud.* **1**, 5–24 (2014).
8. Lahaye, C. *et al. Quat. Geochronol.* **30**, 445–451 (2015).
9. Parenti, F. *Quat. Geochronol.* **30**, 48–53 (2015).
10. Pobiner, B. *Am. Sci.* **104**, 110–117 (2016).
11. Carlisle, E. M. *Ciba Found. Symp.* **121**, 123–139 (1986).
12. Haslam, M. *et al. Curr. Biol.* **26**, R521–R522 (2016).

This article was published online on 19 October 2016.

DNA REPAIR

Telomere-lengthening mechanism revealed

Shortening of the ends of chromosomes limits a cell's lifespan. Some cancer cells avoid this fate through a mechanism called alternative lengthening of telomeres, molecular details of which have now been defined. [SEE ARTICLE P.54](#)

CAITLIN M. ROAKE & STEVEN E. ARTANDI

During cell division, the genome is duplicated by DNA polymerase enzymes. However, each chromosome's ends are incompletely replicated during duplication, because DNA polymerases require an RNA primer 5' to the region being synthesized. This means that the repetitive DNA sequences called telomeres that cap the ends of chromosomes shorten at each division, and this shortening limits the replicative lifespans of most cells. During cancer development, cells acquire the ability to divide indefinitely by circumventing telomere shortening — either by upregulating the enzyme telomerase, which extends telomeres, or by activating a mechanism termed alternative lengthening of telomeres (ALT), which is based on a common method of DNA repair, homologous recombination. On page 54, Dilley *et al.*¹ reveal a

mechanism that underlies ALT and identify an unusual DNA polymerase that mediates this process.

During homologous recombination, a double-strand break in the DNA of one chromosome is repaired by a DNA polymerase using template DNA that is taken from a matching sister chromatid — an identical DNA molecule generated during replication. Cancer cells that use ALT often show higher levels of DNA damage at telomeres than do non-ALT cells², which may predispose them to use homologous recombination to repair breaks in telomeric DNA. In human ALT cancer-cell lines, evidence of enhanced homologous recombination at telomeres includes an increase in telomere exchanges between sister chromatids compared to other cell lines³, and evidence for the copying of DNA tags from one chromosome end to another⁴. An estimated 10–15% of tumours use ALT to maintain their

telomeres, making this process an important target for cancer therapy^{5,6}. However, dissection of the molecular mechanisms that underlie ALT has been challenging.

There are two ways in which telomeres might use homologous recombination to maintain their length. In the first model, unequal exchange of DNA between sister telomeres creates a longer and a shorter telomere, one of which is inherited by each daughter cell. The cells that inherit longer telomeres eventually outcompete those that have shorter telomeres. In the second model, which is increasingly gaining favour, telomeric DNA is synthesized using an existing piece of telomeric template, either from another telomere or from free molecules of repetitive DNA called extrachromosomal telomeric DNA that are found in ALT cells⁵.

Dilley *et al.* provide strong support for the second model. To encourage homologous recombination in ALT cells, the authors exploit a system they had previously engineered⁷ to create targeted double-strand breaks in telomeres by fusing the FokI nuclease enzyme, which cleaves DNA, to the telomere-binding protein TRF1. They observe a tenfold increase in telomeric DNA synthesis after TRF1–FokI induction in cells known to use ALT. Furthermore, they show that this synthesis is unidirectional and processive — capable of synthesizing long tracts of telomere repeats typically 20 kilobases long (the length of an average ALT telomere). The kinetics of this synthesis are consistent with those that could

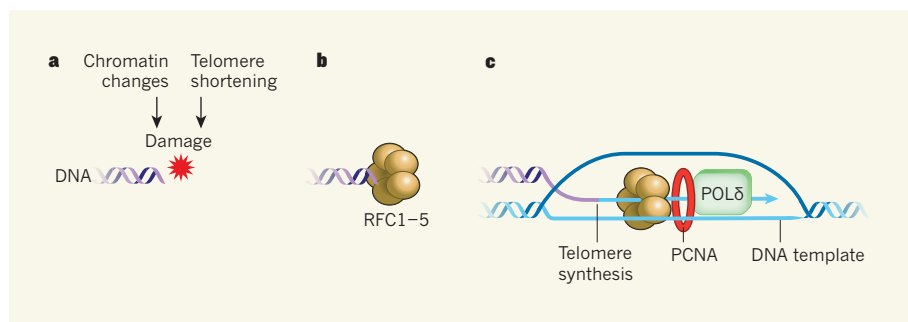


Figure 1 | Break-induced telomere synthesis. Some cancer cells maintain the telomeric DNA sequences that cap the ends of their chromosomes through a mechanism dubbed alternative lengthening of telomeres (ALT). Dilley *et al.*¹ outline a process called break-induced telomere synthesis, which underlies ALT. **a**, Double-stranded DNA breaks can arise in telomeres for several reasons, including telomeric shortening during cell division or changes to the way in which telomeric DNA is packaged as chromatin. **b**, In ALT cells, the protein complex RFC1–5 rapidly binds to double-strand breaks in telomeres. **c**, RFC1–5 recruits the protein PCNA and the DNA polymerase enzyme POL δ to the break. This protein complex synthesizes long tracts of DNA using a complementary strand of DNA as a template from which to replicate the telomere.

cause large fluctuations in telomere length, as is seen in ALT cells⁸.

The characteristics and kinetics of this DNA synthesis match those of a phenomenon called break-induced replication, which is a telomere-maintenance mechanism in yeast strains that lack telomerase⁹. Break-induced replication is a form of homologous recombination that initiates DNA replication when only one end of a double-strand break shares sequence similarity with a template. Dilley *et al.* term this process in mammalian ALT cells break-induced telomere synthesis.

The authors next set out to characterize the proteins responsible for break-induced telomere synthesis. The protein Rad51 has a key role in homologous recombination, and is required for break-induced replication in yeast⁹. But, surprisingly, Dilley and colleagues found that Rad51 was dispensable for break-induced telomere synthesis in ALT cells. Rather, a complex that consists of the polymerase POL δ and the proteins PCNA and RFC1–5 is found at sites of DNA damage in ALT cells and is required for break-induced telomere synthesis (Fig. 1). The authors theorize that this atypical complex is responsible for the dominant pathway of telomere synthesis in ALT cells.

Although Dilley *et al.* shed light on the mechanisms underlying ALT in cancer cells, their findings also open up new questions. For instance, the authors demonstrated that they could trigger break-induced telomere synthesis in both ALT and telomerase-producing cells, so why is this method of telomere replication not operative in most cancer cells? It is unclear what induces the ALT mechanism and how that mechanism is specifically sustained in the 10–15% of cancer cell types that use ALT. The authors provide one possible explanation — that ALT cells have higher rates of persistent telomere damage than other cancer cells.

Alternatively, it might be that there is a change in the way in which telomeric DNA

is packaged around histone proteins to form chromatin. Disruption of histone function has been shown¹⁰ to induce ALT-like characteristics in cells, suggesting a mechanistic link between altered telomere histones and the ALT mechanism. Moreover, mutations in a chromatin-remodelling protein complex, ATRX–DAXX, are highly recurrent in human ALT tumours^{11–14}. The current work does not address how mutations in the ATRX–DAXX complex lead to ALT, but this will be an interesting avenue for further investigation.

Dilley and colleagues' link between break-induced telomere synthesis and ALT provides insights that might help us to

further understand how ALT is initiated and maintained in human cancer cells. In the future, a more in-depth understanding of these processes might lead to the development of therapies targeting human cancers that depend on ALT. ■

Caitlin M. Roake and Steven E. Artandi are in the Departments of Medicine and of Biochemistry and the Cancer Biology Program, Stanford University School of Medicine, Stanford, California 94305, USA. e-mail: sartandi@stanford.edu

1. Dilley, R. L. *et al.* *Nature* **539**, 54–58 (2016).
2. Cesare, A. J. *et al.* *Nature Struct. Mol. Biol.* **16**, 1244–1251 (2009).
3. Londoño-Vallejo, J. A., Der-Sarkissian, H., Cazes, L., Bacchetti, S. & Reddel, R. R. *Cancer Res.* **64**, 2324–2327 (2004).
4. Dunham, M. A., Neumann, A. A., Fasching, C. L. & Reddel, R. R. *Nature Genet.* **26**, 447–450 (2000).
5. Cesare, A. J. & Reddel, R. R. *Nature Rev. Genet.* **11**, 319–330 (2010).
6. Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S. & Reddel, R. R. *EMBO J.* **14**, 4240–4248 (1995).
7. Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. *Cell* **159**, 108–121 (2014).
8. Perrem, K., Colgin, L. M., Neumann, A. A., Yeager, T. R. & Reddel, R. R. *Mol. Cell. Biol.* **21**, 3862–3875 (2001).
9. Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. *Nature* **448**, 820–823 (2007).
10. O'Sullivan, R. J. *et al.* *Nature Struct. Mol. Biol.* **21**, 167–174 (2014).
11. Pickett, H. A. & Reddel, R. R. *Nature Struct. Mol. Biol.* **22**, 875–880 (2015).
12. Heaphy, C. M. *et al.* *Science* **333**, 425 (2011).
13. Lovejoy, C. A. *et al.* *PLoS Genet.* **8**, e1002772 (2012).
14. Schwartzentruber, J. *et al.* *Nature* **482**, 226–231 (2012).

This article was published online on 19 October 2016.

FLUID DYNAMICS

Turbulence in a quantum gas

The discovery of a cascade of sound waves across many wavelengths in an ultracold atomic gas advances our understanding of turbulence in fluids governed by quantum mechanics. [SEE LETTER P.72](#)

BRIAN P. ANDERSON

Microscopic droplets of ultracold atomic gases known as Bose–Einstein condensates might seem among the least-useful substances for studies of turbulence. These droplets typically contain about a billion times fewer atoms than are found in a single human cell, and exist at temperatures of a few billionths of a degree above absolute zero¹. At these temperatures, the motion of the atoms is incredibly slow. Can turbulence really be induced, sustained and

probed in these ultracold atomic gases? Indeed they can, and such turbulence studies are gaining traction, linking atomic and quantum physics with classical fluid dynamics². On page 72, Navon *et al.*³ report observations of a Bose–Einstein condensate as it is driven into a turbulent state and find evidence for a cascade of wave-like excitations, opening up new possibilities for exploring the universality of turbulence.

Chief among the features that are shared by turbulent fluids is a characteristic distribution of kinetic energy between the components of the fluid that have different momenta. To

picture such a distribution, imagine slowly adding cream to coffee and stirring so that the cream is quickly mixed in. The stirring drives energy into low-momentum (long-wavelength) flows, which might be temporarily visible as the cream traces large eddies in the coffee. The liquids are soon mixed and these flows are no longer easily identifiable.

With continued stirring, more energy is injected into low-momentum flows, and the energy initially in these flows is transferred to higher-momentum flows because of the nonlinear dynamics of the fluid. Eventually, energy from high-momentum flows is dissipated as heat because of the viscosity of the coffee. By continuously stirring the coffee, a cascade of excitations can be achieved: energy is injected into the system, transferred through the different momentum components, and finally dissipated (Fig. 1).

Identifying cascades of excitations is a central goal of turbulence studies. These cascades correspond to a power-law dependence of a dynamic quantity such as energy on the wavenumber, a quantity that is proportional to the magnitude of the momentum. Such signatures of turbulence have been experimentally confirmed in countless systems, including superfluid helium^{4,5} (liquid helium held at temperatures so low that it has zero viscosity). Navon and colleagues are the first to identify turbulent cascades of wave-like excitations in an ultracold atomic gas.

The authors' experiment uses a Bose–Einstein condensate (BEC) that consists of about 10^5 rubidium atoms. The atoms are trapped in a 3D cylindrical 'box' about 30 micrometres long, with walls formed by laser light. This trap gives the BEC a uniform density, which ensures that the characteristics of the turbulent cascade are the same throughout the condensate.

Navon *et al.* use a time-dependent magnetic field to shake the cloud of atoms, injecting energy into the low-momentum modes of the BEC. Rather than directly measuring the kinetic-energy spectrum, the authors determine the momentum distribution: the fraction of atoms that have a value of the momentum within any given range. They find that, for short shaking times, the majority of the atoms are in low-momentum modes and few atoms are in high-momentum modes.

After further shaking, interactions between the atoms, which lead to nonlinear dynamics of the BEC, push the atomic population into higher-momentum modes. Continued shaking then replenishes the source of low-momentum excitations, and any atoms that populate high-momentum modes are lost from the trap. Finally, after a total shaking time of about 2 seconds, the authors find that a steady cascade of excitations has been established. This is revealed by the power-law dependence of the momentum distribution on the wavenumber, and represents the

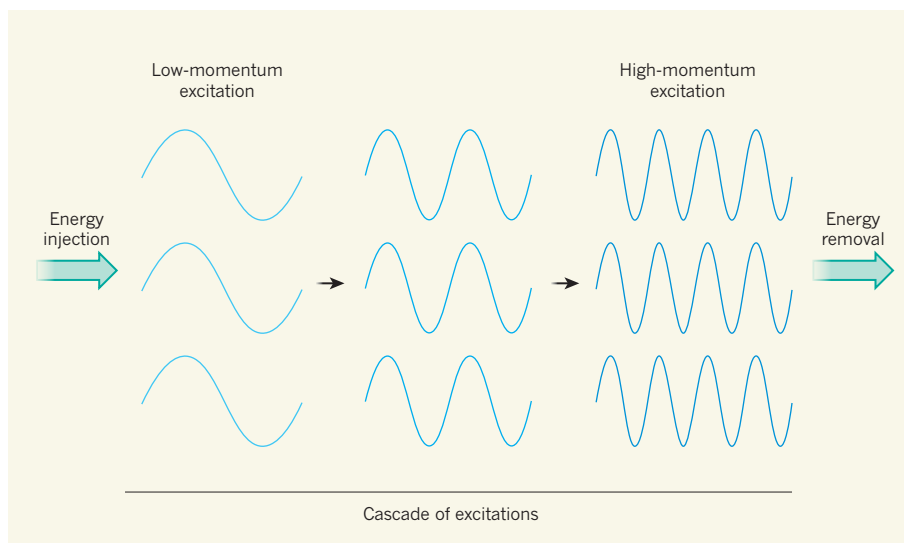


Figure 1 | A turbulent cascade. Navon *et al.*³ find evidence for a cascade of wave-like excitations in an ultracold atomic gas. This simple schematic illustrates how energy in low-momentum (long-wavelength) excitations cascades to high-momentum (short-wavelength) excitations. First, energy is injected into the system, manifesting as excitations of the low-momentum sound modes of the fluid. Because of the fluid's nonlinear dynamics, energy is transferred to successively higher-momentum modes. Eventually, energy is dissipated as heat from high-momentum excitations (in a typical classical fluid), or high-momentum atoms are lost from the system (in the authors' experiment). Navon and colleagues show that excitations covering a wide range of momenta are present simultaneously in their system — a characteristic feature of turbulence.

primary result of the authors' study.

Navon and collaborators' measuring technique involves first suddenly removing the box that confines the atoms, then allowing the atom cloud to expand freely and, finally, acquiring a 2D image of the cloud. For turbulent fluids, the atoms have enough kinetic energy that the interactions between them are comparatively weak. The images of the cloud therefore provide scaled representations of the momentum distributions that existed before the box was removed.

The authors have expertly tackled one aspect of measuring turbulence in ultracold atomic gases, yet many problems remain to be solved before turbulence in these systems can be fully understood. For instance, few theoretical studies have considered analytically the power-law dependence of the momentum distribution for a turbulent atomic fluid. The authors obtain a momentum distribution that is proportional to the wavenumber raised to the power of -3.5 . This is close to the predicted value of -3 for weakly interacting waves in a turbulent compressible superfluid⁶. But the underlying physical mechanisms that give rise to the difference between these values are not understood.

It will also be helpful to extend the cascade over a range of wavenumbers that is larger than that achieved in this experiment: this would allow more flexibility in probing and understanding the nonlinear dynamics of turbulent cascades. Finally, the presence of vortices (localized regions of circulating fluid about a fluid-free core) in this experiment was inferred only through simulations. In general, the amount of energy involved

in vortex excitations compared with sound excitations can greatly affect momentum or energy distributions⁷. Future measurements of vortex dynamics should help researchers to develop a better understanding of turbulence in quantum atomic fluids.

Navon and colleagues' experiment is a crucial step in the further establishment of trapped ultracold atomic gases as systems for studying turbulence. Of broader importance is the contribution of their work to the growing set of techniques for experimentally and theoretically probing turbulence in fluids whose dynamics are governed by quantum mechanics. With substantial theoretical challenges to overcome, the discovery of previously unknown links between turbulence and quantum mechanics is one of the most exciting prospective outcomes of these studies. ■

Brian P. Anderson is at the College of Optical Sciences, University of Arizona, Tucson, Arizona 85721-0094, USA.
e-mail: brian.anderson@optics.arizona.edu

1. Pethick, C. J. & Smith, H. *Bose–Einstein Condensation in Dilute Gases* 2nd edn (Cambridge Univ. Press, 2008).
2. White, A. C., Anderson, B. P. & Bagnato, V. S. *Proc. Natl Acad. Sci. USA* **111** (Suppl. 1), 4719–4726 (2014).
3. Navon, N., Gaunt, A. L., Smith, R. P. & Hadzibabic, Z. *Nature* **539**, 72–75 (2016).
4. Barenghi, C. F., Donnelly, R. J. & Vinen, W. F. *Quantized Vortex Dynamics and Superfluid Turbulence* (Springer, 2001).
5. Barenghi, C. F., Skrbek, L. & Sreenivasan, K. R. *Proc. Natl Acad. Sci. USA* **111** (Suppl. 1), 4647–4652 (2014).
6. Zakharov, V. E., L'Vov, V. S. & Falkovich, G. *Kolmogorov Spectra of Turbulence I* (Springer, 1992).
7. Reeves, M. T., Anderson, B. P. & Bradley, A. S. *Phys. Rev. A* **86**, 053621 (2012).

Phosphate on, rubbish out

A previously unknown way in which cells mark proteins for destruction has been found in bacteria — phosphorylation of the amino acid arginine targets proteins for degradation by protease enzymes. [SEE ARTICLE P.48](#)

ARTI TRIPATHI & SUSAN GOTTESMAN

Bacteria are expert at adapting to changing lifestyles. Dealing with damaged or misfolded proteins is a key challenge during adaptation to stressful conditions such as high temperature. In such situations, chaperone proteins restore unfolded proteins to their functional, folded states and protease enzymes degrade damaged proteins¹. But how are proteins that have been damaged beyond redemption recognized and sent for degradation? On page 48, Trentini *et al.*² show that a surprisingly simple degradation tag — phosphorylation of the amino acid arginine — has a central role in the handling of heat-damaged proteins in the bacterium *Bacillus subtilis*.

Protein degradation in eukaryotes and prokaryotes (organisms with or without a cellular nucleus, respectively) is carried out by protease-enzyme complexes³, which require the nucleotide ATP for their activity. Access to the protease must be controlled to ensure

that only damaged or unwanted proteins are degraded. The protease active site is in a chamber that is accessible only to proteins that have been unfolded and moved there by the ATPase-enzyme portion of the protease-enzyme complex (Fig. 1). Thus, binding and unfolding of proteins by the ATPase governs degradation selectivity.

The selection process for protein degradation often depends on the attachment of molecular 'tags' to the protein to be degraded. For example, eukaryotes might attach the polypeptide ubiquitin to proteins targeted for degradation by proteases⁴. Tagging is also found in prokaryotic archaea and mycobacteria⁵, which use a non-ubiquitin polypeptide tag (Fig. 1).

However, most bacteria select proteins for degradation using other mechanisms than tagging. One such mechanism is the exposure of an amino-acid motif known as a degron⁶, an intrinsic part of a protein. Changes in binding partners of proteins in multi-protein

complexes can hide or reveal degrons. Proteins can also be targeted for degradation if they are bound and specifically delivered to the protease by an adaptor protein, which provides a supporting role in the degradation process, rather than a direct enzymatic one. Modifications or changes in the availability of adaptors and protein-binding partners can determine whether or not a specific protein is degraded at a particular time.

Although the mechanisms that target proteins for destruction are well understood, how damaged proteins are recognized is less clear. When proteins are misfolded as a result of heat or oxidation, they might expose an internal degron that is normally shielded^{1,7}, resulting in the protein's destruction. However, little is known about other possible mechanisms, and Trentini and colleagues' study provides additional mechanistic insights.

McsB, a *B. subtilis* protein that is involved in the heat-shock response⁸, was identified as a protein-kinase enzyme that phosphorylates arginine^{9,10}. McsB-kinase phosphorylation promotes degradation of the protein CtsR (refs 11, 12), which represses expression of heat-shock genes. A previous study¹⁰ had found that more than 100 proteins underwent McsB-dependent phosphorylation, and Trentini and colleagues decided to test whether arginine phosphorylation might cause protein degradation.

The authors investigated degradation by the ClpCP-protease complex in *B. subtilis*. This complex consists of the ClpP protease and ClpC, which is an ATPase and unfoldase enzyme. Under heat-shock conditions, Trentini *et al.* used an inactive version of ClpP to trap proteins undergoing degradation, and, by means of proteomic techniques, identified 20 different trapped proteins that carried a phosphorylated arginine tag; they estimate that 25% of ClpP substrates are marked by arginine phosphorylation.

To demonstrate the role of arginine phosphorylation in degradation, the authors turned to *in vitro* biochemical studies involving purified proteins. Previous investigations of protein degradation by ClpCP found that the process involved adaptor proteins, and McsB was thought to be an adaptor protein¹². Trentini *et al.* found that casein (an unfolded protein commonly used to study protein degradation) could be degraded by ClpCP, but that degradation required either the ClpC adaptor protein MecA or McsB. However, phosphorylated casein was degraded by ClpCP without the action of an adaptor protein, suggesting that the protease recognizes the phosphorylation tag directly. Consistent with a model in which the ATPase contains a binding site for phosphorylated arginine, the authors found that a free molecule of phosphorylated arginine bound to ClpC.

The amino-terminal domain of most members of the bacterial AAA family of unfoldases

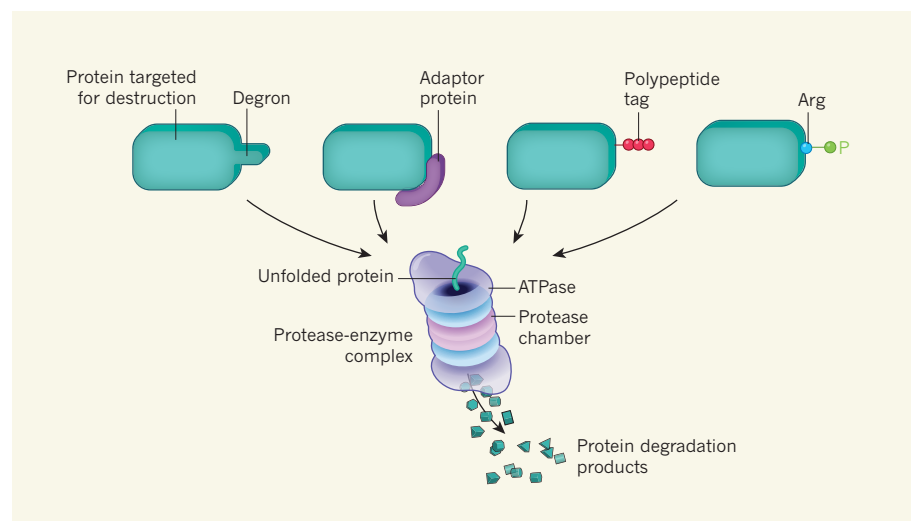


Figure 1 | Protein destruction by protease enzymes. All cells require mechanisms that selectively target proteins for destruction by protease enzymes. Some of the possible mechanisms known to act in prokaryotes (cells without a nucleus) are shown here. A protein can be targeted for destruction if it exposes an amino-acid motif known as a degron, or if the protein is recognized and bound specifically by an adaptor protein or modified through the direct attachment of a polypeptide tag molecule. Trentini *et al.*² have identified a previously unknown type of protein-destruction tag in bacteria — phosphorylation (P) of some of the protein's arginine (Arg) amino acids. Once marked for destruction, a protein transits to the protease-enzyme complex. The protein is recognized and unfolded by an ATPase enzyme and then enters the inner protease chamber, where it is destroyed.

is involved in binding adaptor proteins or proteins to be degraded³. Trentini *et al.* obtained an X-ray crystal structure of the N-terminal domain of ClpC in the presence of a phosphorylated arginine. They identified a binding site for phosphorylated arginine in ClpC that could simultaneously accommodate the positively charged arginine and the negatively charged phosphoryl group of the phosphate.

The binding site in ClpC for phosphorylated arginine partially overlaps with the binding site for the MecA adaptor. Trentini and colleagues mutated two amino acids in the N terminus of ClpC, thereby abolishing its binding of phosphorylated arginine, but still enabling MecA-mediated protein binding. This allowed the authors to test directly whether protease degradation of proteins that have phosphorylated arginines is a necessary part of the heat-shock response. They found that cells lacking ClpC or containing a mutant form of ClpC that does not bind to phosphorylated arginine had a similarly reduced ability to recover from heat-shock stress. This suggests that, during heat shock, ClpC is required to recognize and degrade proteins marked by arginine phosphorylation.

Integrating these findings into our

understanding of heat-shock regulation in bacteria will require more work. It is not clear what changes occur in proteins or McsB at high temperatures that lead to increased arginine phosphorylation. However, McsB and the phosphorylated-arginine binding site in ClpC are evolutionarily conserved across many members of a group known as Gram-positive bacteria, so perhaps arginine phosphorylation might be a widely used mechanism to tag proteins for destruction. Intriguingly, McsB is necessary for *Staphylococcus aureus* to act as a pathogen¹³.

Future experiments should try to unravel the relative importance of the degradation of misfolded proteins compared with the degradation of regulatory proteins during stress survival. Perhaps the machinery for adding and recognizing phosphorylated arginine on proteins will provide new drug-development targets, for example in *S. aureus*.

More broadly, the many biological roles of phosphorylation have now been expanded even further. The ability of this modification to mark proteins for degradation, coupled with the probable perturbation of protein function by the addition of a negatively charged phosphate group to a positively charged arginine,

suggests that Trentini and colleagues' work might be relevant far beyond the *B. subtilis* bacteria they have studied. ■

Arti Tripathi and Susan Gottesman are in the Laboratory of Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892, USA.
e-mail: gottesms@helix.nih.gov

1. Wickner, S., Maurizi, M. R. & Gottesman, S. *Science* **286**, 1888–1893 (1999).
2. Trentini, D. B. *et al.* *Nature* **539**, 48–53 (2016).
3. Olivares, A. O., Baker, T. A. & Sauer, R. T. *Nature Rev. Microbiol.* **14**, 33–44 (2015).
4. Hershko, A. & Ciechanover, A. *Annu. Rev. Biochem.* **67**, 425–479 (1998).
5. Jastrab, J. B. & Darwin, K. H. *Annu. Rev. Microbiol.* **69**, 109–127 (2015).
6. Sauer, R. T. & Baker, T. A. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
7. Neher, S. B., Flynn, J. M., Sauer, R. T. & Baker, T. A. *Genes Dev.* **17**, 1084–1089 (2003).
8. Krüger, E., Zühlke, D., Witt, E., Ludwig, H. & Hecker, M. *EMBO J.* **20**, 852–863 (2001).
9. Fuhrmann, J. *et al.* *Science* **324**, 1323–1327 (2009).
10. Schmidt, A. *et al.* *Mol. Cell. Proteom.* **13**, 537–550 (2014).
11. Elsholz, A. K., Michalik, S., Zühlke, D., Hecker, M. & Gerth, U. *EMBO J.* **29**, 3621–3629 (2010).
12. Kirstein, J., Dougan, D. A., Gerth, U., Hecker, M. & Turgay, K. *EMBO J.* **26**, 2061–2070 (2007).
13. Wozniak, D. J., Tiwari, K. B., Soufan, R. & Jayaswal, R. K. *Microbiology* **158**, 2568–2576 (2012).

LINGUISTICS

Sound and meaning in the world's languages

The sounds of words that represent particular meanings are usually thought to vary arbitrarily across languages. However, a large-scale study of languages finds that some associations between sound and meaning are widespread.

W. TECUMSEH FITCH

Plato's dialogue *Cratylus* begins with a debate. Socrates is asked whether the sounds of words are simply arbitrary conventions, as Hermogenes suggests, or if sounds are reflective in some way of their meaning, as Cratylus proposes. Socrates argues for the latter option, holding that although many words have arbitrary relations to their meanings, 'good' words are distinguished by a correspondence between sense and sound — their sound somehow suits their meaning. However, most researchers today accept the linguist Ferdinand de Saussure's updated version¹ of Hermogenes' viewpoint that the connection between sound and meaning in language is essentially arbitrary, with a few minor exceptions.

Writing in *Proceedings of the National Academy of Sciences*, Blasi *et al.*² address this old debate by exploring an unprecedented number

of languages for indications of associations between sound and meaning, uncovering substantial data that support Socrates' viewpoint. The authors find that, although most words vary arbitrarily, certain associations between speech sounds and the meanings of the words that contain them surface time and again among populations spread across the globe.

There are many categories of sound–meaning correspondence. The most easily recognizable form is termed onomatopoeia, in which the sound of a word corresponds to the sound made by an animal or object. For example, bird names such as cuckoo or chickadee represent phonetic attempts to imitate the characteristic sounds made by these birds. However, even here a certain amount of arbitrariness is present. The crowing of a cock is 'cock-a-doodle-doo' in English but 'kikeriki' in German.

Overlapping somewhat with onomatopoeia are what are known as ideophones. The sound of the entire word of an ideophone has a role

in conveying an iconic meaning that can go beyond simply imitating sounds. Such ideophones are uncommon in English, but one example is the comic-book word 'kapow'. Yet in many other languages, including Korean, Japanese and the African language Ewe, there are large numbers of ideophones, which are frequently used in expressive speech — for example, the Japanese term 'doki doki' expresses the thumping heartbeat of excitement.

Finally, the most pervasive form of sound–meaning correspondence is termed sound symbolism, referring to the situation in which some part of a word or words has a non-arbitrary association between sound and meaning, but other components are arbitrary. A well-known example is the use of different vowels to represent size: the /i/ vowel sound (pronounced 'ee') in the word 'teeny' or 'weeny' symbolizes small size, whereas the /u/ vowel sound (pronounced 'oo') in 'huge' or 'gargantuan' indicates large size. Similarly, in the famous *bouba-kiki* shape–sound effect³, words for sharp, angular objects tend to have consonant sounds such as 'k', whereas smooth, rounded objects are associated with consonants such as 'r' and 'b' (Fig. 1). Both of these phenomena have been extensively studied experimentally and seem to be robust across cultures and age groups^{4–6}.

Other potential examples in English include systematic sound similarities in words possessing related meanings. For example, the sound 'gl-' often indicates shiny visual phenomena (such as glitter or gleam), whereas 'sl-' occurs frequently in words with negative connotations

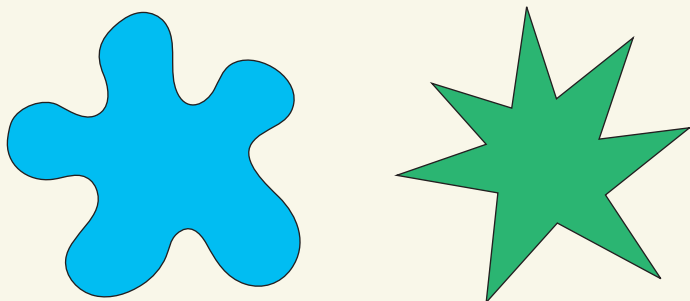


Figure 1 | The bouba–kiki phenomenon. When asked which of the shapes shown are probably named ‘bouba’ or ‘kiki’ in an unfamiliar foreign language, people consistently choose the rounded shape (blue) for *bouba* and associate the sharp-edged shape (green) with *kiki*³ (*bouba* and *kiki* are invented words). This link between the consonants ‘b’ and ‘k’ and round and sharp shapes, respectively, is consistently found in speakers of different languages, and occurs when other invented words are tested²: for example, ‘baluma’ is also associated with the rounded shape. Blasi *et al.*² find that such relationships between the sound and meaning of words are more common and widespread than previously suspected.

(like slime, sludge, slum or slander). The non-arbitrary sound–meaning associations studied by Blasi *et al.* belong in the category of sound symbolism.

Most previous studies of sound symbolism have explored less than 100 languages. Blasi and colleagues started by assembling phonetically transcribed word lists from an enormous variety of languages. They prepared more than 6,000 lists of about 30 basic words and their meanings, as well as conducting a detailed examination of 328 word lists containing 100 items. There is no clear dividing line between a ‘dialect’ and a ‘language’ — linguists often quip that a language is a dialect with an army and a navy. Nonetheless, these lists cover roughly 60% of the world’s languages and 85% of family groups of related languages (such as the Romance language-family grouping that includes Italian, French and Spanish). Although the lists are short, they encompass much of the basic vocabulary in each language and represent the most common nouns and verbs.

Blasi *et al.* then searched these lists for systematic biases in the probability of particular sound–meaning associations (controlling for the biases expected as a result of the overall occurrence of the sounds). They aimed to find robust, widespread sound-symbolic phenomena, and screened out associations found in only one or a few languages. They also aimed to take into account such factors as language relatedness, word length or language-specific constraints.

The result was a surprisingly long list of robust, widespread sound–meaning associations. Reassuringly, Blasi and colleagues found that the vowel /i/ is widely associated with small size, and ‘r’ with roundness, as noted by previous researchers^{4–6}. Other positive associations found by the authors also seem intuitive, such as ‘n’ for nose, ‘t’ for tongue (as in ‘lingual’) and ‘m’ in words for mother or

breasts. Negative associations uncovered included the observation that the sounds ‘b’ and ‘m’ are unlikely to be present in the word for teeth. However, some of the associations identified are particularly hard to fathom, such as ‘s’ for dog (such as *canis* in Latin) or ‘a’ for fish (like *pescado* in Spanish).

Where does this leave the debate between the views of Socrates and de Saussure? As de Saussure argued, most of the sounds in words remain arbitrary, but these data from Blasi and colleagues indicate that matches of sound to sense might be more widespread and pervasive than has been discovered by previous smaller-scale approaches.

The underlying processes driving these

biases between sound and meaning remain a topic for future research. It seems possible that some of these connections reflect the shared sensory and cognitive biology of our species, which over time filters out ‘bad’ words that don’t sound right, and favours words that have a good fit between sense and sound. Over many generations, this filtering will result in selection of variant words and pronunciations, leading to a process of cultural evolution in which words obeying Socrates’ dictum — that good words have sounds that suit their meaning — will have a higher probability of persisting. This hypothesis can be easily tested in laboratory experiments exploring cultural evolution^{7–9}, adding new fuel to the ancient arguments in *Cratylus*. ■

W. Tecumseh Fitch is in the Department of Cognitive Biology, Faculty of Life Sciences, University of Vienna, Vienna 1020, Austria. e-mail: tecumseh.fitch@univie.ac.at

1. de Saussure, F. *Course in General Linguistics* (McGraw-Hill, 1916).
2. Blasi, D. E., Wichmann, S., Hammerström, H., Stadler, P. F. & Christiansen, M. H. *Proc. Natl Acad. Sci. USA* **113**, 10818–10823 (2016).
3. Milan, E., Iborra, O., de Cordoba, M. J., Juárez-Ramos, V., Artacho, M. A. R. & Rubio, J. L. *J. Consc. Stud.* **20**, 84–102 (2013).
4. Hinton, L., Nichols, J. & Ohala, J. (eds) *Sound Symbolism* (Cambridge Univ. Press, 1994).
5. Köhler, W. *Gestalt Psychology* 2nd edn (Liveright, 1947).
6. Sapir, E. *J. Exp. Psychol.* **12**, 225–239 (1929).
7. Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B. & Laland, K. N. *Science* **335**, 1114–1118 (2012).
8. Kirby, S., Cornish, H. & Smith, K. *Proc. Natl Acad. Sci. USA* **105**, 10681–10686 (2008).
9. Mesoudi, A., Whiten, A. & Laland, K. N. *Behav. Brain Sci.* **29**, 329–347; discussion 347–383 (2006).

This article was published online on 26 October 2016.

PARTICLE PHYSICS

Axions exposed

Physicists are hunting for a particle called the axion that could solve two major puzzles in fundamental physics. An ambitious study calculates the expected mass of this particle, which might reshape the experimental searches. SEE LETTER P.69

MARIA PAOLA LOMBARDO

The strong nuclear force binds the constituents of protons and neutrons together and is well described by a theory called quantum chromodynamics (QCD). Within experimental error, the strong force is time-symmetric — the behaviour of strong interactions does not change if the flow of time is reversed. However, the equations of QCD might contain a symmetry-violating term that can theoretically take any value. And nature has chosen to set this term to zero. Why is this? The leading explanation is that the term is not zero, but is cancelled out by the presence

of a ‘neutralizing’ particle called the axion^{1–3}, which has yet to be discovered. Such a particle could also solve a long-standing problem in cosmology by making up dark matter, the ‘missing’ matter in the Universe. On page 69, Borsanyi *et al.*⁴ calculate the expected mass of the axion with unprecedented accuracy, a result that could be useful for understanding the properties of this mysterious particle and directing strategies for its detection⁵.

If axions exist, they would have been produced abundantly during the earliest moments of the Big Bang. Determining the expected mass of this particle therefore requires knowledge of the properties of QCD

at extremely high temperatures. At such temperatures, the axion mass would be time-dependent and proportional to the square root of a quantity called the topological susceptibility. As the Universe expanded and cooled, the mass would have become constant. Consequently, a detailed understanding of both the thermodynamics of QCD (in particular, the temperature dependence of the topological susceptibility) and the expansion of the Universe is necessary to predict the present-day axion mass.

The topological susceptibility can be calculated by exploiting the lattice formulation of QCD. In this approach, space-time is considered to be a four-dimensional grid of discrete points, rather than a continuum. Elementary particles called quarks reside at these points, whereas gluons — the particles that bind quarks together — exist on the links between them.

The next step is to determine how the system of quarks and gluons fluctuates between different ‘topological sectors’. A topological sector is analogous to a ribbon that has a specific number of twists. In a space-time continuum, the ribbon is a closed loop and the number of twists is fixed — fluctuations between topological sectors cannot occur. However, in lattice QCD, the system is discretized. The ribbon is ‘broken’ into many pieces and the number of twists can change — fluctuations between sectors can easily occur. The topological susceptibility can be computed by considering all of the sectors, but this will not be the correct result for continuous space-time.

By contrast, pushing the system towards the continuum limit implies that the fluctuations will become increasingly less likely. In the continuum, the system will be ‘frozen’ in one particular topological sector, and the susceptibility can no longer be calculated. There are further issues that need to be addressed in the calculation, but even at the simplified level of the above discussion, it is clear that studying the topological properties of QCD is a formidable challenge^{6–10}.

Nevertheless, Borsanyi and collaborators compute the topological properties of high-temperature QCD with a remarkable degree of accuracy and sophistication. They achieve this, thanks to an admirable combination of new ideas and an ingenious use of established techniques. The authors’ main technical improvement with respect to other studies^{6–10} is a beautiful strategy to overcome the problems associated with the reduced number of fluctuations between different topological sectors when approaching the continuum limit. Rather than waiting for fluctuations to occur, the authors perform simulations in a fixed sector, estimate the probability of fluctuations between sectors and then reconstruct the topological susceptibility for a wide range of temperatures.

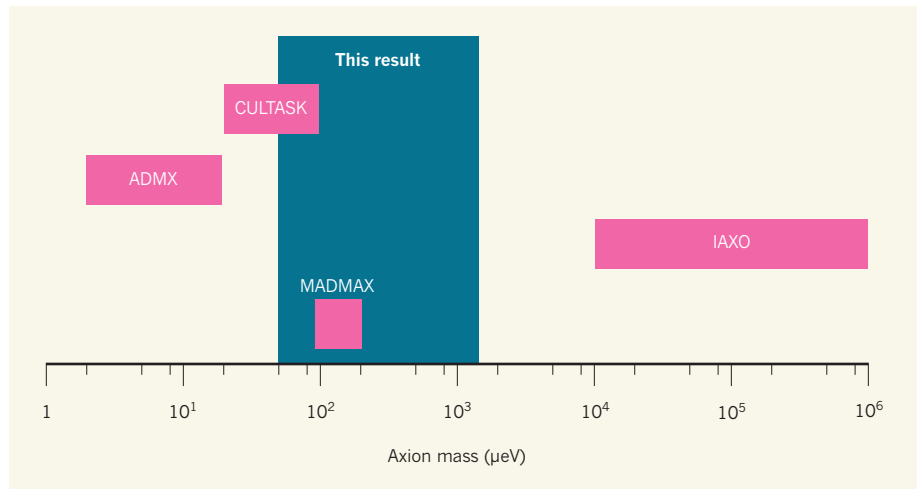


Figure 1 | The hunt for axions. Borsanyi *et al.*⁴ have calculated the expected mass of a hypothetical particle called the axion. Assuming that axions account for dark matter — the ‘missing’ matter in the Universe — the authors determine that the mass of these particles (expressed in energy units based on the electronvolt; eV) must be between 50 and 1,500 μeV . Also shown are estimates of the projected sensitivities of four experiments that have been either proposed or commissioned to look for axions^{4,11}.

Calculating the present-day axion mass requires knowledge of the relationship between the temperature of the Universe and the time elapsed since the Big Bang. Borsanyi and colleagues’ work studies the physics of QCD up to temperatures of more than 10^{15} kelvin, whereas most lattice simulations^{6–10} are limited to temperatures of less than about 7×10^{12} K. The authors discuss the temperatures at which ‘charm’ and ‘bottom’ quarks contribute to the thermodynamics of the system, in addition to the contributions of the other elementary particles. The final results are obtained thanks to a skilful combination of state-of-the-art lattice simulations and other theoretical analyses.

Finally, Borsanyi *et al.* combine their results to obtain a prediction that the axion mass (expressed in energy units based on the electronvolt; eV) lies between 50 and 1,500 μeV (Fig. 1). If the authors’ findings are robust, they will help to guide experimental searches. Current experiments that look for axions¹¹ either try to convert these particles into electromagnetic pulses or observe the effect of axions on the spin of neutrons — in both cases, assumptions need to be made regarding the expected mass.

The accuracy of Borsanyi and colleagues’ results demonstrates the power of their numerical approach and highlights the pivotal role that strong interactions might have in shaping our Universe, by providing a candidate for dark matter. Credit must also be given to other researchers in the field who paved the way for this work by studying lattice topology and its connection to high-temperature QCD. The authors’ work emphasizes that numerical studies need to be triggered and guided by a deep understanding of the underlying physics, but that this can lead to impressive results.

Like experimental data, numerical work should be validated by independent studies, because subtle systematic errors might have a substantial impact on the results. As we have seen, the first step to take when studying strong interactions by numerical means is to discretize their dynamics and then take the system to the continuum limit in a well-controlled way. This is highly complex, and is a subject of research in itself. Different discretization methods need to be explored, and the strategy designed by Borsanyi *et al.* must be further analysed and applied in other, similar work. Let us hope that this remarkable study will motivate researchers to further investigate the properties of the axion, with the aim of giving experiments the robust theoretical input that they need. ■

Maria Paola Lombardo is at the *Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy.*
e-mail: lombardo@lnf.infn.it

1. Peccei, R. D. & Quinn, H. R. *Phys. Rev. Lett.* **38**, 1440–1443 (1977).
2. Weinberg, S. *Phys. Rev. Lett.* **40**, 223–226 (1978).
3. Wilczek, F. *Phys. Rev. Lett.* **40**, 279–282 (1978).
4. Borsanyi, S. *et al. Nature* **539**, 69–71 (2016).
5. Graham, P. W., Irastorza, I. G., Lamoreaux, S. K., Lindner, A. & van Bibber, K. A. *Annu. Rev. Nucl. Part. Sci.* **65**, 485–514 (2015).
6. Berkowitz, E., Buchhoff, M. I. & Rinaldi, E. *Phys. Rev. D* **92**, 034507 (2015).
7. Borsanyi, S. *et al. Phys. Lett. B* **752**, 175–181; preprint at arXiv:1508.06917 (2016).
8. Trunin, A., Burger, F., Ilgenfritz, E.-M., Lombardo, M. P. & Müller-Preussker, M. *J. Phys. Conf. Ser.* **668**, 012123 (2016).
9. Bonati, C. *et al. J. High Energy Phys.* **2016** (3), 155 (2016).
10. Petreczky, P., Schadler, H.-P. & Sharma, S. *J. Phys. Lett. B* (in the press); <http://dx.doi.org/10.1016/j.physletb.2016.09.063> (2016).
11. Patrignani, C. *et al. (Particle Data Group) Chin. Phys. C* **40**, 100001 (2016).

De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure

Jacques-Philippe Colletier^{1*}, Michael R. Sawaya^{2,3*}, Mari Gingery^{2*}, Jose A. Rodriguez², Duilio Cascio², Aaron S. Brewster⁴, Tara Michels-Clark⁴, Robert H. Hice⁵, Nicolas Coquelle¹, Sébastien Boutet⁶, Garth J. Williams⁶, Marc Messerschmidt⁶, Daniel P. DePonte⁶, Raymond G. Sierra⁶, Hartawan Laksmono⁶, Jason E. Koglin⁶, Mark S. Hunter⁶, Hyun-Woo Park^{5,7}, Monarin Uervirojnangkoorn⁸, Dennis K. Bideshi^{5,7}, Axel T. Brunger⁸, Brian A. Federici⁵, Nicholas K. Sauter⁴ & David S. Eisenberg^{2,3}

BinAB is a naturally occurring paracrystalline larvicide distributed worldwide to combat the devastating diseases borne by mosquitoes. These crystals are composed of homologous molecules, BinA and BinB, which play distinct roles in the multi-step intoxication process, transforming from harmless, robust crystals, to soluble protoxin heterodimers, to internalized mature toxin, and finally to toxic oligomeric pores. The small size of the crystals—50 unit cells per edge, on average—has impeded structural characterization by conventional means. Here we report the structure of *Lysinibacillus sphaericus* BinAB solved *de novo* by serial-femtosecond crystallography at an X-ray free-electron laser. The structure reveals tyrosine- and carboxylate-mediated contacts acting as pH switches to release soluble protoxin in the alkaline larval midgut. An enormous heterodimeric interface appears to be responsible for anchoring BinA to receptor-bound BinB for co-internalization. Remarkably, this interface is largely composed of propeptides, suggesting that proteolytic maturation would trigger dissociation of the heterodimer and progression to pore formation.

Of all insects, mosquitoes continue to be the most injurious to human health, spreading devastating diseases such as malaria, filariasis, dengue fever, West Nile encephalitis and, more recently, Zika virus. Synthetic chemical insecticides are a cost-effective means of reducing mosquito vector populations, but their intensive application results in resistance. Most recently, resistance to pyrethroid insecticides has threatened the control of malaria. To deal with the problems of resistance and develop more environmentally sound and sustainable approaches to vector control, various biological and genetic-based control strategies are under development. These include the use of microorganisms such as *wolbachia* endosymbionts that interfere with pathogen development and transmission¹, genetically engineered female mosquitoes incapable of flight or pathogen transmission², and pathogens such as bacteria or fungi engineered for higher efficacy against larvae or adults^{3,4}.

Currently, one of the most environmentally safe and efficient means of controlling mosquitos is the distribution of naturally occurring protein crystals from *Bacillus thuringiensis* subsp. *israelensis* (Bti) and *L. sphaericus* (formerly *Bacillus sphaericus*), as practiced in the United States, Germany, China, Thailand, Ivory Coast and Cameroon⁵. These proteins are toxic to their targets, but harmless to humans and other animals. The specificity and efficiency of the *L. sphaericus* binary toxin, BinAB, presumably arise from its structural complexity, which allows it to navigate through a series of barriers en route and recognize only its intended target. Originating as protoxins inside *L. sphaericus*, BinA and BinB, in a 1:1 ratio, pack into crystals that assemble on the inside of the exosporium, the outside layer of the spore. Formulations of spores containing these crystals are distributed in aquatic environments where mosquitos breed. Larvae ingest the spores,

whereupon the crystals dissolve in the alkaline midgut juices (pH 8–11) and release heterodimers. These are activated by the cleavage of four terminal propeptides. BinB recognizes a glycolipid-anchored maltase receptor located at the microvillar surface of midgut cells⁶ and assists the internalization of BinA, which carries the toxic function⁷. Finally, the BinAB complex enters and kills midgut cells, which results in larval death. Recently, the structure of BinB from *L. sphaericus* was determined⁸. Here, we elucidate the structure of BinAB crystals, revealing features that endow BinA and BinB with their respective functions, and we suggest a mechanism by which alkalinity and proteolytic activation trigger a series of structural rearrangements that navigate BinAB past barriers to reach its target while maintaining the resilient 1:1 association throughout the life cycle⁹.

Determination of the BinAB toxin structure

We collected data from the native BinAB complex and three heavy-atom derivatives using serial femtosecond crystallography (SFX) methods at the Linac Coherent Light Source (LCLS) Coherent X-ray Imaging (CXI) instrument¹⁰, with the aim of phasing by multiple isomorphous replacement with anomalous scattering (MIRAS). We produced nanocrystals of *L. sphaericus* BinAB (Fig. 1a, b) by recombinant expression in *B. thuringiensis* cells and jetted them across the pulsed X-ray free-electron laser (XFEL) beam using either a gas dynamic virtual nozzle (pH 7 crystals)¹¹ or an electrospinning injector (pH 5 and pH 10 crystals)¹² (Extended Data Fig. 1a). Initial structure factor amplitudes were obtained with cctbx.xfel¹³, and then corrected for partiality with cctbx.prime¹⁴ (Extended Data Table 1). This procedure was essential to the success of phasing (Fig. 1c). Heavy-atom sites were successfully located for all three derivatives (Extended Data Fig. 1b–d), and their

¹Institut de Biologie Structurale (IBS), Univ. Grenoble Alpes, CEA, CNRS, 38044 Grenoble, France. ²UCLA-DOE Institute for Genomics and Proteomics, Department of Biological Chemistry, University of California, Los Angeles, California 90095-1570, USA. ³Howard Hughes Medical Institute, University of California, Los Angeles, California 90095-1570, USA. ⁴Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁵Department of Entomology and Graduate Program in Cell, Molecular and Developmental Biology, University of California, Riverside, California 92521, USA. ⁶Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA. ⁷Department of Biological Sciences, California Baptist University, Riverside, California 92504, USA. ⁸Molecular and Cellular Physiology, and Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

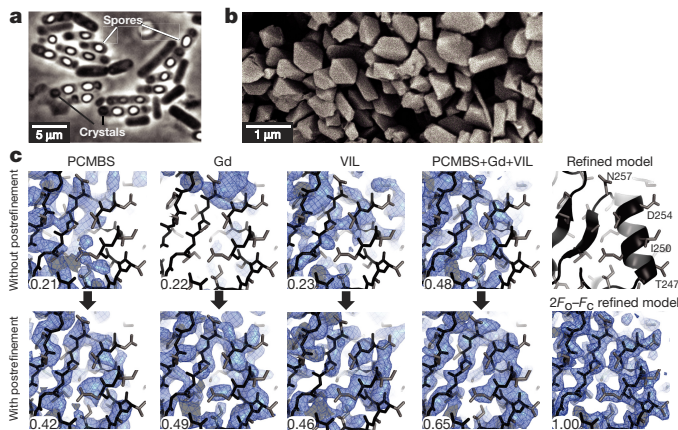


Figure 1 | De novo phasing of SFX data collected from nanocrystals of *L. phaeus* BinAB. **a**, **b**, Phase-contrast micrograph of *B. thuringiensis* sporulated cells engineered to produce BinAB nanocrystals (**a**) and scanning electron micrograph of isolated BinAB nanocrystals (**b**). **c**, The contribution of partiality refinement to the accuracy of the experimental phases is illustrated by maps calculated without (top row) and with (bottom row) the post-refinement procedure¹⁴. The left three columns show SIRAS-phased maps from *p*-chloromercuribenzenesulfonate (PCMBs), Gd, and vaporizing iodine labelling (VIL) derivatives. The fourth column shows MIRAS phased maps from all three derivatives. The right column shows the refined model and $2F_o - F_c$ map. The improvement from post-refinement is quantified by the correlation coefficient (CC) (at the bottom of each panel) between each experimental map and the map phased by the final refined coordinates. Main chain and side chain atoms are shown in black and grey sticks, respectively.

phasing information combined (Extended Data Figs 1e, 2a) to produce a map of sufficient quality that 60% of the BinAB complex could be traced automatically (Extended Data Fig. 2b, c). Subsequent manual building led to a model with an R_{work}/R_{free} of 0.164/0.200 at 2.25 Å resolution (Extended Data Table 1).

The maps revealed two rod-shaped molecules, BinA and BinB, each approximately 100 Å long and 25–30 Å in diameter (Fig. 2a). They resemble each other closely (1.7 Å root mean squared deviation (r.m.s.d.) for 329 pairs of α -carbons; Extended Data Fig. 2d), reflecting the 28% sequence identity (46% similarity) covering nearly their full lengths (Extended Data Fig. 2e). Each subunit comprises two domains: an N-terminal β -trefoil domain (BinA residues 1–155, BinB residues 1–198) and a C-terminal pore-forming domain (PFD) (BinA residues 156–370, BinB residues 199–448).

The most notable structural differences between BinA and BinB are located on the carbohydrate-binding modules of the trefoil domains, suggesting that these may contribute to the distinct roles of BinA and BinB in intoxication. All three modules (α , β and γ) of BinA appear structurally capable of binding carbohydrate; however, BinB contains four loop insertions relative to BinA (residues 62–70, 111–117, 139–143, and 179–185), which distort the trefoil's pseudo-three-fold symmetry (Fig. 2c and Extended Data Figs 2e, 3a). One insertion, residues 62–70, completely obstructs the α -module through a disulfide-linked tether, Cys67–Cys161 (Fig. 2c and Extended Data Fig. 3b). These differences suggest that BinB has a lesser role in carbohydrate binding, or perhaps an adaptation to diverse carbohydrates. Carbohydrate binding has been linked with toxicity. Numerous mutations in these modules reduce toxicity (Supplementary Discussion and Supplementary Tables 1, 2) and sugars such as chitobiose are potent antagonists of the BinA toxin¹⁵. The outward-facing orientation of the modules in the BinAB dimer indicates that they are accessible to cell-surface glycoproteins and glycolipids, which may aid in concentrating BinAB at the cell surface before receptor binding (Fig. 3b).

Relatively few structural features distinguish BinA from BinB in the PFD, suggesting that pore assembly may be heteromeric¹⁶. The PFD topology is characteristic of the aerolysin family of pore-forming toxins, comprising a 60 Å-long antiparallel β -sheet, folded into a sandwich at one end (sandwich subdomain), and adherent to a putative membrane-spanning segment (transmembrane subdomain)¹⁷ at the other end (sheet subdomain; Fig. 2a and Extended Data Fig. 4).

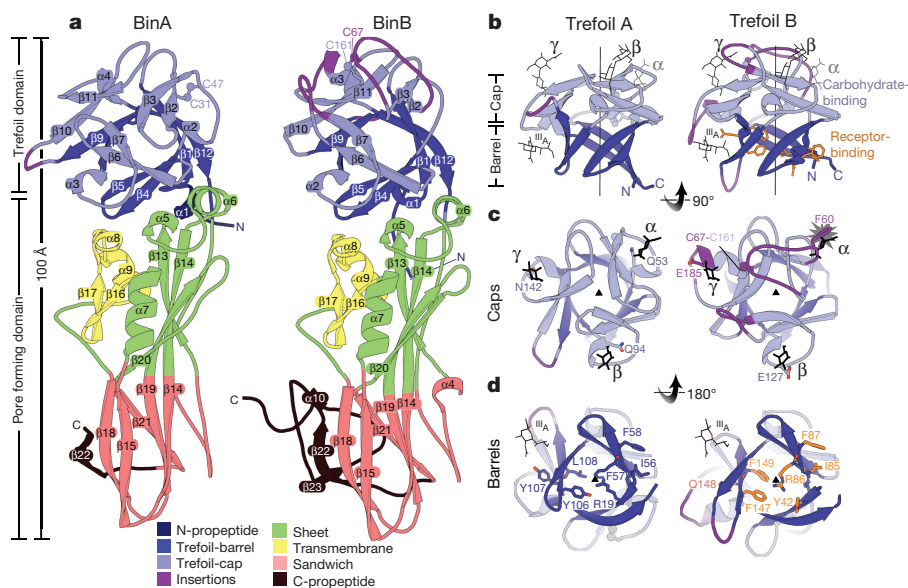


Figure 2 | BinA and BinB folds and carbohydrate-binding modules. **a**, BinA and BinB are structurally similar to each other, each composed of trefoil and pore-forming domains. The most noticeable differences correspond to insertions in surface loops on the trefoil domains (purple). **b**, The trefoil domains are composed of barrel and cap subdomains. In the cap subdomain, there are three canonical carbohydrate modules (α , β and γ). Carbohydrate molecules superimposed from structures of haemagglutinin (3AH1)²⁹ and haemolytic lectin (1W3G)³⁰ are shown in black sticks. These occupy the α , β , and γ binding modules. A fourth

carbohydrate-binding site, marked III_A, is a minor site observed in haemagglutinin²⁹. **c**, View of the cap subdomains along the pseudo-three-fold symmetry axis; loop insertions (purple) break the symmetry in BinB. The starburst indicates a steric overlap between the modelled carbohydrate and the α -module of BinB. The conflict arises from the 9-residue insertion in this loop, tethered by a disulfide bond, C67–C161. **d**, View of the barrel subdomains along the pseudo-three-fold symmetry axis. BinB residues implicated in receptor binding are shown in sticks (orange). Structurally analogous residues are shown on BinA.

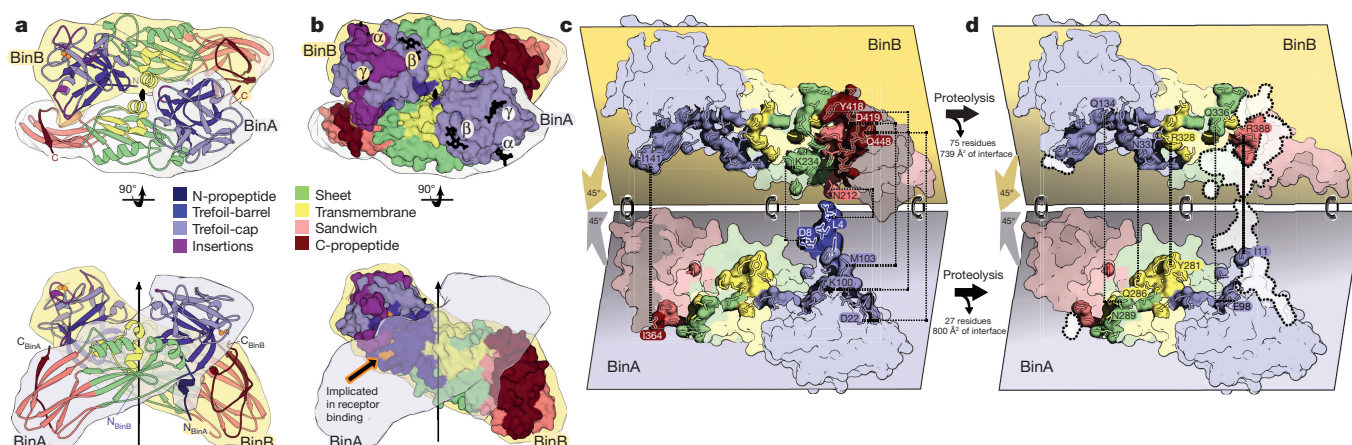


Figure 3 | BinAB dimer assembly is weakened by proteolysis. **a**, The BinAB dimer shown here manifests the largest intermolecular interface in the crystal. The vertical arrow and lens-shaped symbol (black) indicate the position of the pseudo-two-fold rotation axis that relates BinA and BinB. The interface extends over all four domains in both molecules. Propeptides (dark blue and dark red) play a substantial role in the interface. **b**, Surface representation of the BinAB dimer shows that the canonical carbohydrate modules are accessible, but the receptor-binding epitope is not, implying that carbohydrate binding occurs first and then a conformational change exposes the receptor binding epitope. **c–d**, Transformation of the BinAB

This structural homology suggests that BinAB may form pores by a mechanism similar to aerolysin¹⁸, in which the transmembrane subdomain refolds into a hairpin, oligomerizes into a β -barrel, and inserts into the membrane. However, no experiment has revealed whether these pores would be homo- or hetero-oligomeric. We find that the degree of structural similarity between the PFDs of BinA and BinB (1.4 Å r.m.s.d. over 203 α -carbons; Extended Data Fig. 2d) is even closer than that observed between LukF and Hlg2 (1.5 Å r.m.s.d. over 241 α -carbons), which adopt alternate positions around the ring-shaped heterooctameric pore of γ -haemolysin¹⁹. A heteromeric assembly analogous to γ -haemolysin might explain why BinA and BinB form pores more efficiently in combination than either component separately^{16,20}.

Propeptides at the BinAB dimer interface

The crystal packing reveals a unique and exceptionally large heterodimer interface, specifying a resilient 1:1 association between the two components. In this interface, the 100 Å-long BinA and BinB molecules cross each other at a 30° angle, forming an X shape (Fig. 3a, b). This interface is so large (burying 1,855 Å² on BinA and 1,800 Å² on BinB) that it accounts for nearly half of all the intermolecular interface area in the crystal. It exceeds by far the threshold value (856 Å² per monomer) estimated to discriminate between biological and artificial dimers²¹. It is three times larger than the next largest interface in the crystal, and therefore it is likely to be the only one of the total seven intermolecular interfaces preserved after dissolution of the crystals (Extended Data Figs 5, 6 and Supplementary Tables 3–9). Its shape complementarity (0.65) is similar to that observed in antibody–antigen interfaces²² and electrostatic complementarity is evident (Extended Data Fig. 7a, b). The fit and extent of this heterodimer association implies stability, which probably contributes to attaining the 1:1 molar ratio shown to be optimal for receptor binding and toxicity (Supplementary Tables 1, 2).

Further examination of the heterodimer interface reveals that its size would be severely reduced by proteolytic activation and that this loss would appear to threaten access of BinA into the cell. Remarkably, 42% of the heterodimer interface involves propeptide segments, that is, 1,539 Å² combined interface surface area (Fig. 3c, d, Extended Data Figs 5c, 6a, c, d, f, and Supplementary Tables 3, 10). Therefore, removal of these cleavable segments before internalization would dismantle what may be the essential interface tethering BinA to its chaperone, the cell-surface-receptor-bound BinB. Indeed, deletion of the largest of

interface accompanying proteolysis. The BinAB dimer is split apart to reveal the interface. All four subdomains and three propeptides contribute to the BinAB interface (c). Dashed lines connect specific propeptide residues in contact across the dimer interface. Dissociation of the propeptides following proteolysis eliminates 42% of the interface (d). Dotted lines encompassing white patches mark the interface lost after dissociation of the propeptides. Dashed lines connect select residues remaining in contact across the dimer interface following propeptide dissociation. The transmembrane subdomain (yellow) is the only subdomain that does not lose contacts after proteolysis.

the four propeptides (BinB residues 396–448) reduces the ability of this truncated mutant to direct the regional binding of BinA to target cells²³. However, we note that owing to the large surface area it buries in BinAB (3,790 Å²), this 53-residue propeptide would be slow to release. Slow release may delay heterodimer dissociation until after internalization, when, at the targeted location, it signals transformation into a pore.

Alkaline-triggered release of BinAB dimers

Of all the transformations undergone by BinAB over its life cycle, our data bear most directly on the pH-signalled transformation from crystal to soluble dimers. Using crystallography as a means of structure elucidation, we can see in detail the molecular contacts that are the focus of this transformation and monitor their perturbation with elevation of pH from 7 to 10. The most notable attribute of this transformation is the seemingly contradictory combination of both stability and sensitivity that evolved to adapt the crystal to different stages of the life cycle. Crystal stability preserves and stores potency in harsh environments before ingestion, yet alkalinity readily dissolves the crystal in the larval midgut (pH 8–11) after ingestion, releasing BinAB to access cell-surface receptors and activating proteases. The mechanism by which alkalinity lowers the high barrier to crystal dissolution is evident at three levels: amino acid composition, and local, and global structural changes.

Alkalinity may facilitate crystal dissolution through the concerted deprotonation of BinAB's 49 tyrosine hydroxyl groups, occurring around pH 10 ($pK_a^{\text{Tyr}} = 10.1$) (Extended Data Fig. 7c). The frequency of tyrosine residues (6%) is almost twice the average for proteins (3.5%)²⁴. We calculate that pH elevation from 7 to 10 would change the net charge on BinAB from -13.9 to $-73.5e$, thereby destabilizing the crystals through negative electrostatic repulsion (Extended Data Fig. 7d, e). An analogous mechanism was proposed for the dissolution of the ultra-stable viral spindle crystals, which also contain an unusually high frequency of tyrosines (8.6%), some located strategically at crystal contacts²⁵.

Alkalinity observably perturbs four regions in BinAB crystals, identified by peaks in a difference Fourier map ($F_o - F_c$) computed between data sets collected at pH 7 and pH 10 (Fig. 4, Extended Data Fig. 8 and Supplementary Tables 11, 12). These regions involve contacts between molecules or subdomains. Two of these perturbations increase accessibility of propeptide segments to proteases. The N-terminal

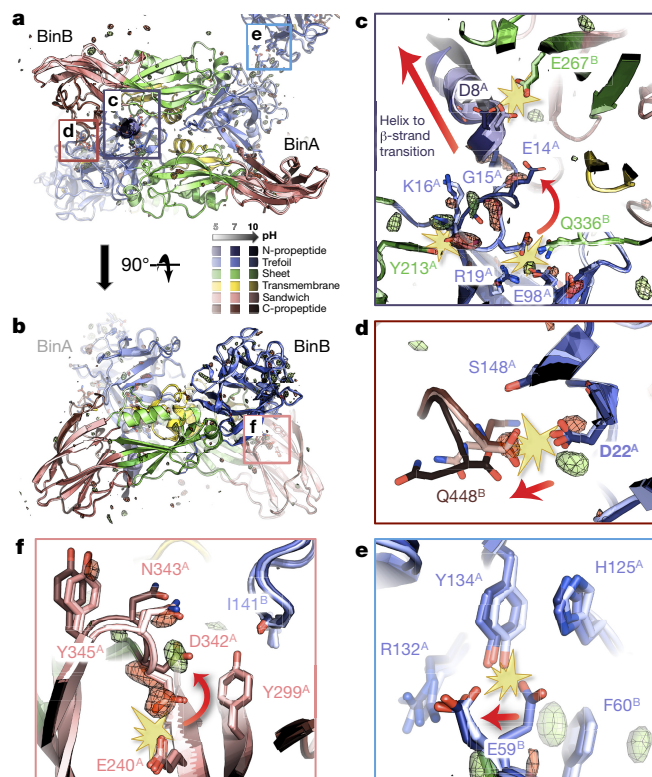


Figure 4 | pH sensing in BinAB crystals. a–f, BinA and BinB cartoons are coloured by subdomain. The pH 5, pH 7 and pH 10 structures are shaded light, medium, and dark, respectively. Symmetry-related molecules are coloured similarly. The $F_o^{[pH10]} - F_o^{[pH7]}$ map is superimposed at $\pm 3.5\sigma$, with positive and negative peaks shown in green and red, respectively. a and b show orthogonal views of the BinAB dimer. The $F_o - F_o$ map reveals four regions (c–f) of the BinAB crystal that are perturbed by elevated pH, probably reflecting early events in crystal dissolution. Rupture of hydrogen bonds and conformational changes are highlighted by starbursts and arrows, respectively. c, Deprotonation of Asp8, Glu14 and Tyr213 triggers a helix to extended β -strand conformational change in the BinA N-terminal propeptide, increasing its accessibility to proteolytic cleavage. d, Deprotonation of the BinB Gln448 terminal carboxylate breaks its H-bond with BinA Asp22, weakening the BinA–BinB dimer interface and making the C-terminal propeptide of BinB available for proteolysis. e, Deprotonation of BinA Tyr134 and His125 results in the rupture of their H bonds with BinB Glu59. Loss of these contacts directly weakens the lattice. f, Negative electrostatic repulsion pushes Asp342 away from Glu240 in BinA. This rearrangement could be an early step in the transformation into a pore, owing to Asp342's location at the junction of the three PFD subdomains.

propeptide of BinA (residues 1–10) unravels from a helix to an extended conformation, following the loss of a hydrogen bond between Gly15(O) and Tyr213(OH) of BinA (Fig. 4c and Extended Data Fig. 9). Similarly, a pair of hydrogen bonds is broken in the BinAB dimer interface between the C-terminal carboxylate of the BinB propeptide (Gln448) and BinA Asp22 side-chain carboxylate (Fig. 4d). The third region is a lattice contact outside the dimer interface, between BinA Tyr134(OH) and BinB Glu59(OE2) which breaks at elevated pH (Fig. 4e and Supplementary Table 7). The fourth region involves loss of a hydrogen bond between BinA Asp342(OD1) and Glu240(OE1). Located at the junction between the transmembrane and sandwich subdomains, this break might be an early step towards pore formation (Fig. 4f and Extended Data Fig. 8). In each of these four regions, deprotonation is presumably the cause of hydrogen-bond disruption, involving either a tyrosine hydroxyl or a carboxylate group paired with an obligate hydrogen bond acceptor (Fig. 4 and Supplementary Discussion).

Alkalinity also induces global hinge motions, potentially straining the crystal lattice and thereby contributing to its dissolution. The trefoils of BinA and BinB move about 0.5 Å closer to their respective PFDs (Extended Data Fig. 9 and Supplementary Discussion). These motions might also foreshadow a rearrangement of the dimer interface, to expose the receptor-binding motif, which is otherwise buried in the dimer interface (Fig. 3a, b).

To validate our model of alkaline-induced crystal dissolution and address concerns that the structural changes we attribute to pH elevation may have instead originated from unintended radiation damage (Supplementary Discussion), we mutated one of the four pH-sensitive switches that we identified, BinA Asp22. Recall that upon pH elevation, withdrawal of donor hydrogen atoms breaks the bifurcated hydrogen bond between carboxylates of Asp22 and the C terminus of BinB (Fig. 4d). Mutation from Asp to Asn should reduce toxicity by decreasing sensitivity of this interaction to alkaline-induced rupture and thereby delay release of the C terminus for proteolytic activation. This mutation did not change the appearance of the crystals; however, it reduced crystal solubility by 30% in the period 30–90 min after pH elevation (Extended Data Fig. 7f) and increased LC₅₀ and LC₉₅ by 11.6-fold and 24-fold, respectively (Supplementary Table 13). The effect of this single conservative substitution in a 93-kDa complex provides convincing validation of our model.

Our structure of BinAB illuminates several important molecular events in the life cycle of the toxin. These include discovery of (1) four pH-sensitive switches that facilitate crystal dissolution in the larval midgut; (2) a large heterodimer interface that explains how heterodimers persist after dissolution; (3) three competent carbohydrate-binding modules in BinA that may assist in directing heterodimers to the cell surface; and (4) the potential to disrupt the heterodimer interface by proteolytic activation, thereby signalling remodelling. Our success in *de novo* MIRAS phasing of a crystallographic asymmetric unit nearly three times larger than any previously phased *de novo* by SFX^{26–28}, and from crystals approximately 50 unit cells per edge, suggests this approach could be applied again in other cases where crystal size is limiting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January; accepted 7 September 2016.

Published online 28 September 2016.

- Moreira, L. A. *et al.* A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium*. *Cell* **139**, 1268–1278 (2009).
- Alphey, N., Bonsall, M. B. & Alphey, L. Combining pest control and resistance management: synergy of engineered insects with Bt crops. *J. Econ. Entomol.* **102**, 717–732 (2009).
- Federici, B. A., Park, H.-W. & Bideshi, D. K. Overview of the basic biology of *Bacillus thuringiensis* with emphasis on genetic engineering of bacterial larvicides for mosquito control. *Open Toxicology J.* **3**, 83–100 (2010).
- Fang, W. *et al.* Development of transgenic fungi that kill human malaria parasites in mosquitoes. *Science* **331**, 1074–1077 (2011).
- Worrall, E. & Fillinger, U. Large-scale use of mosquito larval source management for malaria control in Africa: a cost analysis. *Malar. J.* **10**, 338 (2011).
- Darbois, I., Nielsen-LeRoux, C., Charles, J.-F. & Pauron, D. The receptor of *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae) midgut: molecular cloning and expression. *Insect Biochem. Mol. Biol.* **31**, 981–990 (2001).
- Berry, C. *et al.* Genetic determinants of host ranges of *Bacillus sphaericus* mosquito larvicidal toxins. *J. Bacteriol.* **175**, 510–518 (1993).
- Srisucharitpanit, K. *et al.* Crystal structure of BinB: a receptor binding component of the binary toxin from *Lysinibacillus sphaericus*. *Proteins* **82**, 2703–2712 (2014).
- Opota, O. *et al.* *Bacillus sphaericus* binary toxin elicits host cell autophagy as a response to intoxication. *PLoS One* **6**, e14682 (2011).
- Liang, M. *et al.* The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. *J. Synchrotron Radiat.* **22**, 514–519 (2015).
- DePonte, D. P. *et al.* SEM imaging of liquid jets. *Micron* **40**, 507–509 (2009).
- Sierra, R. G. *et al.* Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1584–1587 (2012).

13. Hattne, J. *et al.* Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nat. Methods* **11**, 545–548 (2014).
14. Uervirojnangkoorn, M. *et al.* Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *eLife* **4**, 05421 (2015).
15. Broadwell, A. H. & Baumann, P. Proteolysis in the gut of mosquito larvae results in further activation of the *Bacillus sphaericus* toxin. *Appl. Environ. Microbiol.* **53**, 1333–1337 (1987).
16. Cokmus, C., Davidson, E. W. & Cooper, K. Electrophysiological effects of *Bacillus sphaericus* binary toxin on cultured mosquito cells. *J. Invertebr. Pathol.* **69**, 197–204 (1997).
17. Melton, J. A., Parker, M. W., Rossjohn, J., Buckley, J. T. & Tweten, R. K. The identification and structure of the membrane-spanning domain of the *Clostridium septicum* alpha toxin. *J. Biol. Chem.* **279**, 14315–14322 (2004).
18. Degiacomi, M. T. *et al.* Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat. Chem. Biol.* **9**, 623–629 (2013).
19. Yamashita, K. *et al.* Crystal structure of the octameric pore of staphylococcal γ -hemolysin reveals the β -barrel pore formation mechanism by two components. *Proc. Natl Acad. Sci. USA* **108**, 17314–17319 (2011).
20. Schwartz, J. L. *et al.* Permeabilization of model lipid membranes by *Bacillus sphaericus* mosquitocidal binary toxin and its individual components. *J. Membr. Biol.* **184**, 171–183 (2001).
21. Pongstingl, H., Henrick, K. & Thornton, J. M. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins* **41**, 47–57 (2000).
22. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
23. Oei, C., Hindley, J. & Berry, C. Binding of purified *Bacillus sphaericus* binary toxin and its deletion derivatives to *Culex quinquefasciatus* gut: elucidation of functional binding domains. *J. Gen. Microbiol.* **138**, 1515–1526 (1992).
24. Klapper, M. H. The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem. Biophys. Res. Commun.* **78**, 1018–1024 (1977).
25. Chiu, E. *et al.* Structural basis for the enhancement of virulence by viral spindles and their *in vivo* crystallization. *Proc. Natl Acad. Sci. USA* **112**, 3973–3978 (2015). 10.1073/pnas.1418798112
26. Barends, T. R. M. *et al.* De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **505**, 244–247 (2014).
27. Yamashita, K. *et al.* An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Sci. Rep.* **5**, 14017 (2015).
28. Fukuda, Y. *et al.* Redox-coupled proton transfer mechanism in nitrite reductase revealed by femtosecond crystallography. *Proc. Natl Acad. Sci. USA* **113**, 2928–2933 (2016).
29. Nakamura, T. *et al.* Sugar-binding sites of the HA1 subcomponent of *Clostridium botulinum* type C progenitor toxin. *J. Mol. Biol.* **376**, 854–867 (2008).
30. Mancheño, J. M., Tateno, H., Goldstein, I. J., Martínez-Ripoll, M. & Hermoso, J. A. Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J. Biol. Chem.* **280**, 17251–17259 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the help of the following people during data collection: S. Lee, J. Koralek, R. Shoeman, S. Botha, B. Doak and O. Zeldin. We thank A. Volveda for advice regarding sequence-wise Fourier difference map integration; J. Brooks-Bartlett and E. Garman for help with dose calculations; and M. Weik for discussions and continuing support. We thank the HCIA program of HHMI, the W.M. Keck Foundation (grant 2843398), the NIH (grant AG-029430), National Science Foundation (grant MCB 0958111) and DOE (DE-FC02-02ER63421) (to D.S.E.), the France Alzheimer Foundation (FA-AAP-2013-65-101349) and the Agence Nationale de la Recherche (ANR-12-BS07-0008-03) (to J.-P.C.), NIH grants GM095887 and GM102520 for data-processing methods (to N.K.S.), and NIH grant AI45817 (to B.A.F.). Support by the CNRS (PEPS-SASLELX-2013, PEPS-SASLELX-2014) funded travel to LCLS. Use of the LCLS at SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, and Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The CXI instrument was funded by the Linac Coherent Light Source Ultrafast Science Instruments project, itself funded by the DOE Office of Basic Energy Sciences. Parts of the sample injector used at LCLS for this research were funded by the National Institutes of Health, P41GM103393, formerly P41RR001209.

Author Contributions J.-P.C., M.R.S., M.G., J.A.R., D.C., B.A.F. and D.S.E. designed and coordinated the project. M.G. carried out the *in vivo* production of BinAB nano-crystals. H.-W.P., D.K.B. and B.A.F. engineered bacterial strains to produce enlarged BinAB crystals. R.H.H. and H.-W.P. performed toxicity assays. R.H.H. designed the mutagenesis protocol. D.K.B. performed crystal solubilization assays. M.R.S. performed heavy atom derivatizations. J.-P.C., M.R.S., J.A.R., D.C., A.S.B., T.M.-C., S.B., G.J.W., M.M., D.P.D., R.G.S., H.L., J.E.K., M.S.H., N.C., M.U. and N.K.S. acquired and processed data. J.-P.C. and M.R.S. carried out the MIRAS and MR phasing, and built and refined the atomic models. R.G.S. and H.L. developed the MESH-on-a-stick injector. J.-P.C., M.R.S., and D.S.E. prepared the manuscript with input from M.G., J.A.R., D.C., A.S.B., T.M.-C., R.G.S., M.S.H., A.T.B., B.A.F., and N.K.S.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 5FOY (pH 7 structure), 5FOZ (pH 10 structure) and 5G37 (pH 5 structure). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-P.C. (colletier@ibs.fr) and D.S.E. (david@mbi.ucla.edu).

Reviewer Information Nature thanks C. Berry, H. Chapman and P. Wang for their contribution to the peer review of this work.

METHODS

In vivo production of *Lysinibacillus sphaericus* BinAB crystals. *L. sphaericus* BinAB crystals were grown in an acrySTALLiferous *Bacillus thuringiensis* strain engineered to improve crystal expression, 4Q7 (*Bacillus* Stock Center at Ohio State University, Columbus, Ohio), which allowed growth of much larger crystals than occur naturally in *L. sphaericus*. Testing for mycoplasma was not carried out. There is no evidence that growth of the *L. sphaericus* BinAB crystals in *B. thuringiensis* in any way affected the structure of BinAB; the toxicity per unit mass of the Bin crystals produced in *B. thuringiensis* was the same as that of crystals produced in *L. sphaericus*. Assuming that the cytosolic compositions of *L. sphaericus* and *B. thuringiensis* cells do not strongly differ, crystal packing interactions are expected to persist and crystals formed by BinAB in the two types of cells should be identical in terms of space group and unit cell constants. Five hundred millilitres of glucose-yeast-salts (GYS) growth medium (0.1% glucose, 0.2% yeast extract, 0.05% K_2HPO_4 , 0.2% $(NH_4)_2SO_4$, 0.002% $MgSO_4$, 0.005% $MnSO_4$, and 0.008% $CaCl_2$) supplemented with 25 μ g/ml erythromycin was sterilized in a 2-l baffled flask and inoculated with spores from a lyophilized 5-day lysate of *B. thuringiensis* subsp. *israelensis* strain 4Q7 containing plasmid pPHSP-1 (Bti4Q7/pPHSP-1), which encodes BinA and BinB from *L. sphaericus* strain 2362 (ref. 31). Of all known *L. sphaericus* strains, this strain produces the most potent toxin. Cultures were harvested after growth for 5 days at 30 °C with shaking at 250 rpm, monitored by phase contrast light microscopy until sporulation and cell lysis were observed.

We attempted to obtain crystals of BinAB toxin incorporating the unnatural amino acid 3-iodo-tyrosine (3iTyr) at tyrosine positions in BinA and BinB. By exposing early-sporulating cells to excess 3iTyr, phenylalanine, and tryptophan, we aimed to block aromatic amino acid biosynthesis and force the utilization of the iodine-labelled amino acid. Bti4Q7/pPHSP-1 cells were grown to late log phase and exposed to 3iTyr, phenylalanine, and tryptophan at concentrations of 100 mM each. Two further additions of the three amino acids, each increasing the concentration by 65 mM, were made at 24-h intervals. Cultures were harvested after 4 days. However, SDS-PAGE analysis showed partial or no incorporation of 3iTyr in the BinAB crystals. Furthermore, there was no evidence of incorporation in the final structure of BinAB. Following culture lysis, spores, crystals, cells, and cell debris were pelleted by centrifugation at 6,000g for 30 min, then resuspended in 50 ml sterile water. The concentrated lysate was sonicated for 3 min on ice (1 s on, 1 s off (6 min elapsed time); 60% intensity) to lyse the remaining cells. The sonicated lysate was pelleted at 6,000g for 30 min at 4 °C. All subsequent steps were performed at 4 °C. The pellet was washed in 50 ml cold water to remove soluble material and some of the spores, then re-pelleted, and resuspended in 15 ml water. Crystals were isolated from the suspension by sucrose gradient centrifugation as described in ref. 32. Before injecting the sample into the beam, large particles were removed by filtering the material with a 10- μ m stainless steel frit.

Construction of pPHSP1 D22N and pBUSP-1 and pBUSP-1 D22N clones. The recombinant plasmid pPHSP-1, containing BinA and BinB open reading frames under control of the pSTAB promoter³³, was used as a template to amplify two DNA fragments each containing the D22N mutation in the BinA ORF, using Q5 polymerase (New England Biolabs). The first fragment began at the *Mlu*I site near the 3' end of the BinB ORF and went through the D22 region of the BinA ORF, and contained the BinA D22N mutation; this fragment was amplified using primer pairs [1]. The second fragment began at the D22 region and continued to the end of the Bin operon terminator, and was amplified with primer pair [2]. Both fragments had approximately 15-bp overlapping homology regions with the vector on one end and the second fragment on the other. The two fragments were assembled along with pPHSP-1 linearized with *Mlu*I and *Pst*I by recombining using the Choo-Choo Cloning Kit (MCLAB). The entire Bin operon of the resulting clone was sequenced to confirm the presence of the D22N mutation. Full-length fragments of the entire Bin operon of either wild-type pPHSP-1 or pPHSP-1-D22N, including approximately 15-bp vector homology regions on both ends, were amplified using primers [3], and assembled by recombining in pBU4³⁴ linearized with *Kpn*I and *Sal*I. The resulting plasmids, pBUSP-1 and pBUSP-1 D22N, which also contained the 20-kDa helper protein gene as previously described^{35,36} were sequenced and used to transform *Bacillus thuringiensis* strain 4Q7 by electroporation³³.

Primer pairs:

[1] TTCACACTAAAACGCGTTAATGGTGAAATTG (BinA D22N For 1) and CTGCTATTATAAAATTCATAACGCGAATGTACTTTCCTTCTG (BinA D22N Rev 1)

[2] CGTTATGAATTTTATAATAGCGAGTATCCTTTCTGTATACATGCACC (Bin D22N Frag 2 F2 New) and CCAAGCTTGCATGCCTGCAGC (BinA D22N Rev 2)

[3] GTGAATTCGAGCTCGGTACCGAATTCTATTTTCGATTTCAAATTTTC CAAAC (HSP Promoter For) and GGGTGTTAACGTCGACAAACAACA GTTTACATTCTGA (HSP Bin Term. Rev)

Bioassay for mosquito larvicide activity. Lyophilized cultures containing spores and parasporal bodies of the *L. sphaericus* BinAB and BinA^{D22N}B strains were resuspended in ddH₂O. Suspensions were diluted to 6–7 different concentrations, ranging from 0.5 ng/ml to 1 μ g/ml, in 6 oz cups in a final volume of 100 ml. Bioassays were replicated three times using 30 fourth-instars of S-Laboratory (Bin-sensitive) strains of *Cx. quinquefasciatus*. After 24 h of exposure at 28 °C, dead larvae were counted and the 50% and 95% lethal concentrations, respectively (LC₅₀ and LC₉₅), were calculated by Probit analysis (POLO-PC; LeOra Software).

Assay for alkaline-induced crystal dissolution for wild-type BinA–BinB and BinA(D22N)–BinB. The strains producing BinAB (4Q7/pPHSP-1) and BinAB(D22N) (4Q7/pBU-D22N) were grown in nutrient broth + glucose (NBG) supplemented with, respectively, erythromycin (25 mg/ml) and tetracycline (3 mg/ml), for 5 days until >95% of cells had sporulated and lysed. To isolate crystalline inclusions, spore/crystal mixtures collected from 50-ml cultures were resuspended in 15 ml ddH₂O and sonicated twice at 50% duty cycle for 15 s using the Ultrasonic Homogenizer 4710 (Cole-Parmer Instrument Co.). Five-millilitre samples were loaded onto a sucrose gradient cushion (30–65% w/v), which was then centrifuged at 20,000g for 45 min at 20 °C in a Beckman L7-55 ultracentrifuge using the SW28 rotor. Bands containing inclusions were collected and washed three times in ddH₂O, followed by centrifugation at 6,500g for 15 min at 4 °C after each wash, then lyophilized for storage at –20 °C until use. For solubilization assays, 200 μ g or 800 μ g of each crystal preparation was resuspended in, respectively, 100 μ l of 10 mM Tris-Cl (pH 7; baseline control) or 400 μ l of 10 mM Tris-Cl (pH 10; solubilization buffer). Assays were performed in triplicate. For each sample, three 25- μ l aliquots were collected and spun at 16,300g for 2 min to pellet undissolved crystals, and supernatants were transferred to fresh microfuge tubes containing 75 μ l of 10 mM Tris-Cl (pH 10). Absorbance at 280 nm (A₂₈₀) was recorded at various time intervals.

Preparation of the mercury derivative. We produced the mercury derivative by soaking the BinAB crystals in 1 mM *p*-chloromercuribenzenesulfonate (PCMBS) for five months. Specifically, we added 200 μ l of 0.1 M PCMBS to 15 ml crystal slurry at room temperature, and then we passed the mixture through a stainless steel frit with 10- μ m pores. We estimated the concentration of crystals in the slurry by pelleting the crystals in a centrifuge, then noting the crystals occupied approximately 25 μ l of the 1-ml sample volume. The five-month incubation was not intentional. We intended to use the derivative on the day it was prepared, which was during a previous data collection trip; however, this sample immediately clogged the gas dynamic virtual nozzle (GDVN) injector upon its first use. So, virtually none of this large volume of sample passed through the injector. In retrospect, we think the clogging was the result of mercuric iodide precipitation induced by residual iodide ion in the delivery lines; a derivative containing 0.5 M potassium iodide had indeed been run just previously to injecting the PCMBS derivative. To reduce the chance of this same PCMBS sample clogging again, we removed the non-covalently bound PCMBS by washing the crystals twice in deionized water a week before the LCLS experiment. The crystals used for derivatization in this experiment have the same origin as those we used to produce our effectively 'native' data set. They were prepared under conditions intended to substitute 3iTyr in place of tyrosine, but the substitution was not efficient and so they were effectively native before derivatizing with PCMBS.

Preparation of the gadolinium derivative. We produced the gadolinium derivative by soaking the BinAB crystals in 5 mM GdCl₃. More specifically, we added 50 μ l of 0.1 M GdCl₃ to 1 ml crystal slurry at room temperature one day before the diffraction experiment. We estimated the concentration of crystals in the slurry by pelleting the crystals in a centrifuge, then noting the crystals occupied approximately 25 μ l of the 1-ml sample volume. In order to conserve the amount of crystals used in screening for a heavy atom derivative, we recycled the crystal sample that had been used in previous diffraction experiments performed in helium vapour atmosphere (that is, not vacuum). For example, the crystals used in this experiment had been soaking for five months in KI and CsCl since the time they cycled through a GDVN nozzle for data collection in a previous XFEL experiment. The sample could be recycled without fear that the previous experiment had damaged the crystals since over 99.9% of the crystals never intercepted the XFEL beam during the previous XFEL experiment. We washed the KI and CsCl from the crystals by pelleting the crystals and replacing the supernatant with water three times before adding GdCl₃.

Preparation of the VIL derivative. We produced an iodinated tyrosine derivative by adapting the vaporizing iodine labelling (VIL) methods described in ref. 37. The method uses gaseous iodine vapour to derivatize the ortho positions of

accessible tyrosine residues. In our case, rather than derivatizing a few crystals in a standard 1–5- μ l drop, we derivatized 1 ml crystal slurry. To improve accessibility of the iodine vapour to this relatively larger volume of crystals, we distributed the 1 ml of BinAB crystal slurry evenly over nine wells of a glass depression dish (that is, 111 μ l per well). Four 10- μ l aliquots of a KI/I₂ mixture (0.67 M/0.47 M concentration) were placed on the dish, adjacent to but not contacting the crystal slurry in the wells. The plate was sealed inside a sandwich box with 5 ml of water to act as a reservoir. The three components (drops of crystal slurry, the iodine, and the water reservoir) were separated in space but in vapour contact. The sandwich plate was sealed with tape at room temperature. The drops of crystal slurry turned yellow after 3.5 h. The crystals were harvested after 21 h. We estimated the concentration of crystals in the slurry by pelleting the crystals in a centrifuge, then noting the crystals occupied approximately 25 μ l of the 1-ml sample volume. The crystals used in this experiment had the same history of origin as those used for the gadolinium derivative. They had been soaking for five months in KI and CsCl for a previous XFEL experiment, then jetted across the beam in helium vapour atmosphere, and washed three times to remove the non-covalently bound heavy atoms before derivatizing with iodine.

Sample injection using GDVN (pH 7 crystals). BinAB nanocrystals of the native and derivatives (average volume is $0.75 \times 0.5 \times 0.25 \mu\text{m}^3$) were injected using a gas-focused dynamic virtual nozzle (GDVN) liquid microinjection system³⁸ at the Coherent X-ray Imaging (CXI) instrument of LCLS¹⁰. Data were collected on two occasions, one for native data and one for heavy atom data, for a combined time of 245.9 min, with the detector framing at 120 Hz.

Preparation of pH 5 BinAB crystals. Crystals of native BinAB at pH 7 were pelleted by centrifugation at 10,000g for 5 min and then resuspended in pH 5 buffer (0.1 M N-cyclohexyl-3-aminopropanesulfonate (CAPS) pH 10.0, 30% glycerol (v/v), 10% PEG 2000 MME, and 0.1 M NaCl). This preparation was originally intended to be at pH 10; however, it was later discovered that one of the components, polyethylene glycol monomethyl ether 2000 (PEG 2000 MME) was acidic (probably due to age), overpowering the pH 10 CAPS buffer, and bringing the final pH to 5. The pH of the crystal slurry was indicated by colour change on pH paper. The crystals were soaked at pH 5 for 10 min before the diffraction experiment.

Preparation of pH 10 BinAB crystals. Crystals of native BinAB at pH 7 were pelleted by centrifugation at 10,000g for 5 min and then resuspended in pH 10 buffer (0.11 M CAPS pH 10.0, 33% glycerol (v/v), 11% methylpentanediol, and 0.11 M NaCl). The pH of the crystal suspension was verified by colour change on pH paper. The crystals were soaked at pH 10.0 for 5 h before the diffraction experiment.

Sample injection via MESH-on-a-stick (pH 5 and pH 10 crystals). Crystals of native BinAB in pH 5 and pH 10 buffers were delivered in the microfluidic electrokinetic sample holder (MESH) method, described more fully in ref. 12. The MESH injection system was modified (Extended Data Fig. 1a) to interface more readily with the CXI nozzle rods that mount interchangeably into the liquid sample delivery system. The crystal slurries were prepared as described above. A continuous 1.5-m-long fused silica, polyamide-coated capillary of 100 μ m inner diameter and 360 μ m outer diameter was used to deliver the sample into the CXI vacuum chamber. Approximately 800 μ l of sample slurry with glycerol additive was placed in a microcentrifuge tube and placed in a small pressurized sample-holder. The capillary and platinum wire were fed through the pressure cell and immersed in the slurry. A small backing pressure of 5 psig nitrogen gas was applied to aid the injection. The voltage was applied by a Stanford Research Systems PS350 high-voltage source and was held between 4,300 and 4,500 V (currents <1 μ A) while the counter electrode was grounded. The flow rate was not directly measured, but we estimate that the sample consumption was approximately 2 μ l/min, as judged from crude measurements of leftover sample volume. The pH 10 structure presented here was collected in less than 1 h of continuous beamtime and consumed less than 1 ml sample. The sample injection compared favourably to prior attempts at data collection by giving comparable resolution with only hundreds of microlitres of sample consumed.

Data collection, indexing, merging and post-refinement. The sample chamber was at room temperature, under vacuum. Native and derivative pH 7 data sets were collected with an XFEL beam focused to a 1- μ m FWHM spot, and characterized by a wavelength of 1.21 Å (native crystals; 1.9×10^{11} photons per pulse) or 1.41 Å (heavy atom soaked crystals; 7.2×10^{11} photons per pulse). We chose a single wavelength for all three derivatives as a compromise of beam stability, time efficiency, and measurable I^σ for the elements used (Extended Data Fig. 1b). The pH 5 and pH 10 data sets were collected with a XFEL beam focused to a 100 nm FWHM spot (6.4×10^{11} photons per pulse), and characterized by a wavelength of 1.46 Å. Thus, the pH 5 and pH 10 data sets were collected with a ~500-fold higher dose than the pH 7 data set (Extended Data Table 1).

Images were reduced using cctbx.xfel^{13,39}. No thresholding was applied to filter out blank shots; we attempted to index every frame. Furthermore, for the heavy atom-soaked data, if a frame indexed successfully, we removed the primary lattice from the list of candidate reflections and attempted to index again, searching for a secondary lattice. Eight per cent of the final number of indexed images from the heavy atom data was from secondary lattices. In total, 24.5% of all recorded frames from the heavy atom data contained indexable patterns (13.2% for the native data), for a final total of 398,971 indexed patterns across the three derivatives and the native data set collected at pH 7. The data sets collected at pH 5 and pH 10, respectively, consisted of 17,099 and 27,792 indexed patterns. First we merged and post-refined our data without negative intensities included using cctbx.prime¹⁴, which determines relative scale factors for each image, along with partiality estimates for each structure factor measurement. Post-refinement in this manner was critical for multiple isomorphous replacement with anomalous scattering (MIRAS) phasing as described below. Figure 1c illustrates, for the three derivatives and their combination thereof, the extent to which cctbx.prime improved the quality of the intensities, leading to improved phases, and thereby rendering the path of the peptidic chain visible in electron density map, where only broken density was visible before. Resolution cut-offs were determined based on completeness (>90%), redundancy (>4) and CC_{1/2} (>0.14); in all data sets, <I/sigI> in the highest resolution shells was greater than 6. For all data sets, postrefinement corrected for the inherent partiality of SFX data and thus improved the σ -weighted (variance-weighted) average estimate of structure factor amplitudes. Handling of negative intensities was then implemented in cctbx.xfel, allowing the production of meliorated post-refined data sets with intensity distributions that no longer have abnormal reflection intensity metrics (L-test and NZ-test). We used the same criteria for the resolution cut-off, based on completeness on (>90%), redundancy (>4) and CC_{1/2} (>0.14); in all data sets, the resulting <I/sigI> in the highest resolution shells was greater than 0.5. While inclusion of negative intensities dramatically affects <I/sigI>, it results in stabilization of the refinement for the three native structures (pH 5, pH 7 and pH 10), requiring us to impose a lower weight on geometry in phenix.refine to obtain a structure with a plateauing R_{free} . Inclusion of negative intensities in the data sets also resulted in a decrease of noise levels in q -weighted structure factor amplitude Fourier difference maps (see below). At the atomic level, the only noticeable change between structures refined with and without negative intensities included is an increase in the B -factors, reflecting the increase of the Wilson-B—and possibly giving a more realistic description of the BinAB structures at room temperature.

Structure determination by MIRAS. Heavy-atom sites were successfully located for all three derivatives using SHELXD⁴⁰ (Extended Data Fig. 1c, d). The individual derivatives demonstrated limited phasing power, and the maps they produced were not interpretable. Fortunately, the three derivatives were isomorphous (Extended Data Fig. 1e), and so could be combined to obtain a map of sufficient quality to automatically trace a partial BinAB model using the Phenix suite⁴¹ (Fig. 1c and Extended Data Fig. 2a, b), and extend using Arp/Warp⁴². The remaining residues were built manually, leading to a model with $R_{\text{work}}/R_{\text{free}}$ of 0.164/0.200 at a resolution of 2.25 Å (Extended Data Table 1). After each cycle of model rebuilding, reciprocal space refinement (including refinement of coordinates, atomic displacement parameters, TLS parameters and occupancy) was carried out using phenix.refine. The final MIRAS model features residues 1–357 of BinA and 28–448 of BinB. Figure 1c illustrates the benefit of post-refinement to the accuracy of experimental phasing. For this figure, all SIRAS and MIRAS maps were calculated using MLPHARE, followed by ten cycles of density modification with DM⁴³ enforcing 58% solvent content, and displayed at 2.8 Å resolution. The CC quantifies the improvement gained from post-refinement. It reports on the whole asymmetric unit, not just the region illustrated in the figure.

Details of MIRAS phasing. We attempted to determine heavy-atom substructures for each of the three derivatives by supplying both isomorphous and anomalous difference signals (SIRAS), or only anomalous difference signal (SAD) to the program SHELXD⁴⁰ in super-sharpening mode (PSMF = −4). Success was evidenced by the clear separation between populations of false and true solutions ranked by goodness-of-fit. Super-sharpening enhanced this separation. We found that SAD substructure determination was successful for the PCBMS (mercury) and Gd (gadolinium) derivatives, but not for the VIL (iodide) derivative (Extended Data Fig. 1d). SIRAS substructure determinations were, however, successful for all three derivatives (Extended Data Fig. 1d). SIRAS phases of the mercury and iodide derivatives are of better quality than their SAD counterparts, as evidenced by a sharper separation between populations of false and true solutions ranked by goodness-of-fit (Extended Data Fig. 1d). In contrast, the SAD phases of the gadolinium derivative are better quality than its SIRAS phases.

However, for each individual derivative, the correct choice of hand remained ambiguous after initial density modification of SIRAS phased maps and chain

tracing with SHELXD, as judged by a lack of separation in map quality statistics between the two hands, such as contrast, connectivity, and correlation coefficient for auto-traced chains (Extended Data Fig. 2a). We sought to maximize the distinction between hands by systematically computing a series of maps at different estimated solvent contents, from 55 to 73%, and comparing the map quality statistics from opposing hands (Extended Data Fig. 2a). The true solvent content of BinAB crystals is 59.7%. We found that for each value of solvent content tested between 55% and 65%, a consistent choice of hand was indicated by the correlation coefficient of the automatically traced chain. However, at higher solvent content values tested, the opposite hand was indicated. The correlation coefficients of traces (CCs) were below 10% in all cases, and the differences between the CCs of the two hands were in the 0.5–2.5% range. So, the distinction between hands appeared to remain ambiguous at this point. For each derivative, we chose the hand having the most consistently (repeatedly) better correlation coefficient over the range of solvent content values tested. The refined heavy atom sites were input to phenix.autosol from the Phenix⁴¹ software suite for phase refinement, density modification and initial model building.

Attempts to solve the structure based on any of the single derivative (including single isomorphous replacement, single anomalous dispersion and single isomorphous replacement with anomalous scattering) were all unsuccessful. The combination of phase information from the three derivatives nonetheless produced usable maps, allowing us to obtain an experimental BinAB model. Our previous inability to distinguish the correct hand for the heavy atom substructure for the individual derivatives was overcome in Phenix through use of cross difference Fourier techniques, an option that becomes available with multiple derivatives (Extended Data Fig. 2a). Briefly, phenix.autosol (using Solve for phase combination)⁴⁴ was used to test various phase combinations schemes and to refine phases. Although quality maps were obtained by combining anomalous and isomorphous signals from all three derivatives (MIRAS), the best map was obtained combining the anomalous signal of the gadolinium derivative with the combined anomalous and isomorphous differences of the mercury and iodide derivatives. Initial figures of merits (FOMs) were 0.12, 0.14 and 0.19 for the PCMBs, Gd and VIL substructures, respectively. The FOM of the best phase-combined map was 0.19 before density modification in Resolve⁴⁴, and 0.26 after density modification. The phase-and-build routine of Phenix was able to trace 296 residues from this map, resulting in a model with R_{free} and R_{factor} of 0.41 and 0.39, respectively (Extended Data Fig. 2b). Note that we had to optimize the parameters of phenix.autosol to obtain this model; in particular, we specified that the data were weak, that we needed to perform a thorough search, and that we needed to use 'extreme-dm'. The FOM of the density-modified map obtained after running this script was 0.72. From this map, phenix.autobuild was in 30 cycles able to reconstruct 501 residues, of which 183 were placed, corresponding to R_{free} and R_{factor} of 0.35 and 0.32, respectively.

This model and the corresponding phases were then input to a first round of Arp/Warp⁴², fitting 62 additional residues into the BinA and BinB models (518 residues, 245 placed) (Extended Data Fig. 2b). This experimental but incomplete BinAB model displayed R_{free} and R_{factor} of 0.32 and 0.29, respectively. The MIRAS model then underwent another cycle of model building in phenix.autobuild and then a second round of Arp/Warp, resulting in a model with 630 residues, of which 450 were putatively in sequence. This automatically fitted model was, however, unsatisfactory in terms of R_{free} , R_{factor} and goodness of fit of the placed residues in the experimental map. Manual rebuilding was necessary, in order to correct wrongly assigned residues and to fit the complete BinAB sequence into the electron density maps. In one round of manual fitting, an additional 92 residues were built, leading to a total of 722 residues (299 and 423 residues in BinA and BinB, respectively) (Fig. 1c and Extended Data Fig. 2c). A cycle of reciprocal space refinement and yet another cycle of manual rebuilding produced a map which allowed us to fit the 59 missing residues of BinA. The final model displays R_{free} and R_{factor} of 0.200 and 0.164, respectively (Extended Data Table 1), with 96.5% of residues in favoured regions of the Ramachandran plot (0.5% Ramachandran outliers, 1.2% rotamer outliers). Of note, we could also phase and build the structure using MIRAS phases of all derivatives, that is, without excluding the isomorphous differences from Gd, but the initial model turned out to be less complete, further highlighting the higher quality of the SAD phases of the gadolinium derivative, as compared to SIRAS (Extended Data Fig. 2b).

Determination of pH 7 structure by molecular replacement. Molecular replacement for BinAB (strain 2362) was performed with Phaser⁴⁵ using as a starting model the activated BinB structure⁸ (PDB ID: 3WA1) from *L. sphaericus* strain 2297. BinB from the two strains differ in only five amino acids (A104S, R267K, L314Y, F317L, L389M). This search model became available only after we had begun our search for heavy atom derivatives. BinB shares only 28% identity with BinA. Two copies of BinB were found (corresponding to a solvent content of 68%), one of which indeed corresponded to BinB, but the other to BinA. Corresponding

residues were mutated and fit manually into the electron density maps using Coot⁴⁶. After each cycle of model rebuilding, reciprocal space refinement (including refinement of coordinates and atomic displacement parameters) was carried out using phenix.refine⁴¹. The last cycle of refinement was performed using Buster⁴⁷ and included TLS refinement. The R_{work} and R_{free} of the final model were 15.8 and 20.3%, respectively, at a resolution of 2.25 Å. Note that residues 1–11 of BinA and residues 1–34 of BinB were disordered in the electron density map and therefore not included in the model. This model is nearly identical to the independently built and refined MIRAS model (0.102 Å r.m.s.d. over 773 aligned α -carbons; see Supplementary Table 14). The BinAB model obtained by molecular replacement was never used in building the BinAB model to the experimental map. In fact, the two models were built independently by different authors. The similarity between the independently obtained models provided additional assurance of model accuracy.

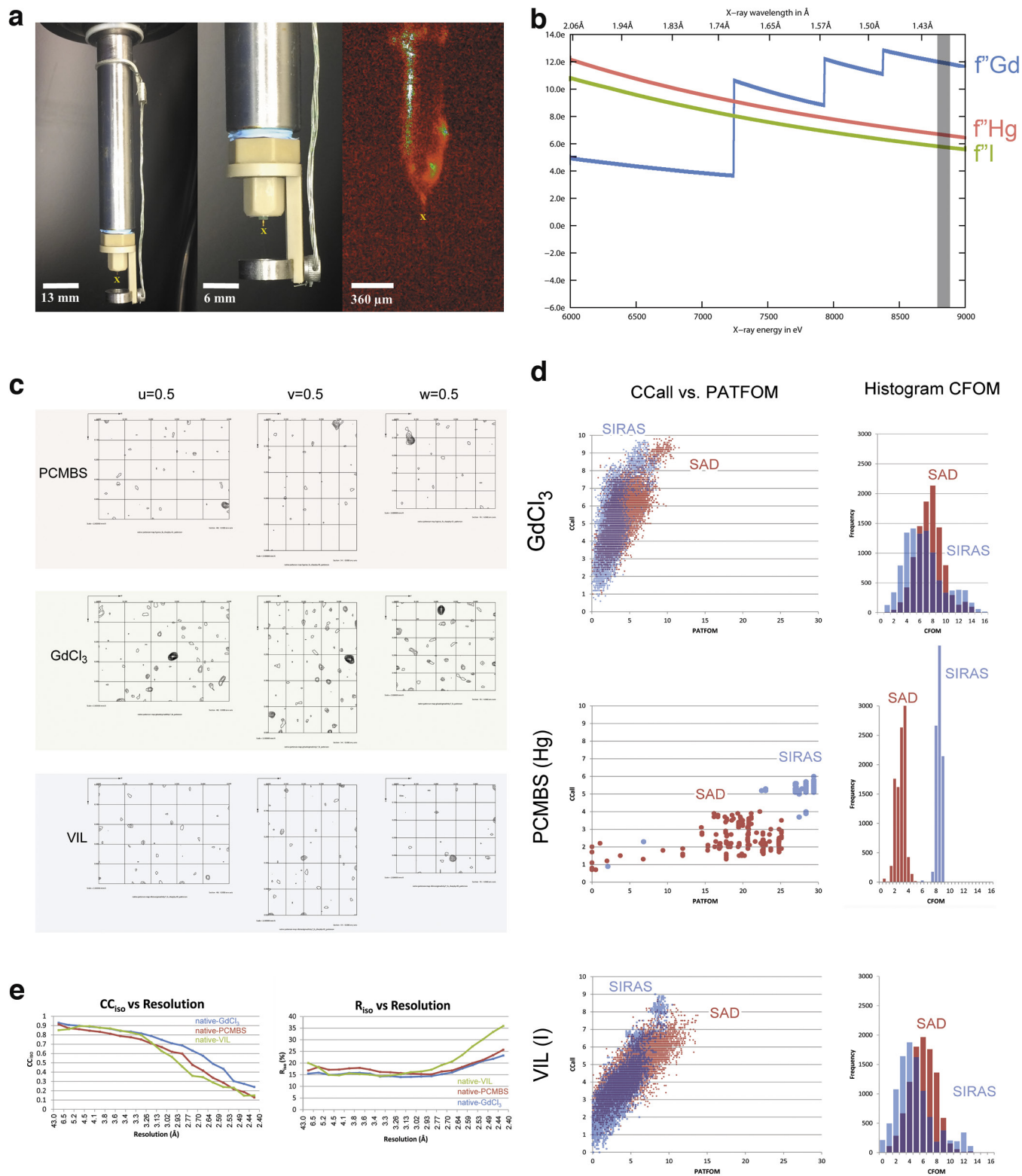
Structure determination of BinAB at pH 5 and 10. We obtained experimental insights into pH-induced conformational changes by calculating structure factor amplitude Fourier difference maps ($F_o - F_c$) between the pH 5, pH 7 and pH 10 data sets. To improve the estimate of structure factor amplitude differences, $F_o - F_c$ maps were q -weighted as described⁴⁸ and produced using a CNS⁴⁹ custom-written script⁵⁰. Application of the q -weighting scheme to the diffraction data sets was essential to eliminate noise and amplify the difference signal. Figure 4 shows the $F_o^{\text{pH10}} - F_o^{\text{pH7}}$ map ($R_{\text{iso}} = 0.28$) phased with the pH 7 model (φ^{pH7}) and highlights four regions that are highly sensitive to pH elevation (Fig. 4c–f). Extended Data Fig. 8 shows, for each of these regions (Extended Data Fig. 8a–d) and for the two BinAB disulfides (Extended Data Fig. 8e, f), the six possible $F_o^{\text{pHj}} - F_o^{\text{pHi}}$, φ^{pHi} maps calculated from the pH 5, pH 7 and pH 10 data sets and structures. Supplementary Tables 11 and 12, respectively, list all BinA and BinB residues that feature peaks higher than $\pm 3.5\sigma$ and larger than 2 voxels in the $F_o^{\text{pH10}} - F_o^{\text{pH7}}$, φ^{pH7} map. Sequence-wise integrations of the $F_o^{\text{pH10}} - F_o^{\text{pH7}}$, φ^{pH7} and $F_o^{\text{pH5}} - F_o^{\text{pH7}}$, φ^{pH7} maps around BinA and BinB are given in Extended Data Fig. 8g and h, respectively.

We phased the pH 5 and pH 10 data sets by molecular replacement with Phaser⁴⁵ using as a starting model the refined MIRAS (pH 7) structure. Conformational changes with respect to the latter were modelled manually. Reciprocal space refinement was performed using phenix.refine and included positional, B -factor, TLS and occupancy refinements. R_{work} and R_{free} of the final pH 5 model are 0.211 and 0.262, respectively, at a resolution of 2.5 Å, with 97.6% of residues in favoured region of the Ramachandran plot (0.4% Ramachandran outliers, 1.5% rotamer outliers). R_{work} and R_{free} of the final pH 10 model are 0.165 and 0.211, respectively, at a resolution of 2.4 Å, with 98.9% of residues in favoured region of the Ramachandran plot (0.0% Ramachandran outliers, 0.0% rotamer outliers). Difference distance matrices (DDMs) are presented in Extended Data Fig. 9. $F_o - F_c$ map integration and DDM calculations were performed using custom-written scripts.

Electrostatic potential maps. pK_a values were assigned using PROPKA⁵¹ and structures protonated in the AMBER force field at pH 7.5 and pH 10.5 using PDB2PQR⁵². Figures were produced using PyMOL⁵³.

- Park, H. W., Ge, B., Bauer, L. S. & Federici, B. A. Optimization of Cry3A yields in *Bacillus thuringiensis* by use of sporulation-dependent promoters in combination with the STAB-SD mRNA sequence. *Appl. Environ. Microbiol.* **64**, 3932–3938 (1998).
- Sawaya, M. R. *et al.* Protein crystal structure obtained at 2.9 Å resolution from injecting bacterial cells into an X-ray free-electron laser beam. *Proc. Natl Acad. Sci. USA* **111**, 12769–12774 (2014).
- Park, H.-W. *et al.* Recombinant larvicidal bacteria with markedly improved efficacy against culex vectors of West Nile virus. *Am. J. Trop. Med. Hyg.* **72**, 732–738 (2005).
- Bourgoignie, C., Delcluse, A., de la Torre, F. & Szulmajster, J. Transfer of the toxin protein genes of *Bacillus sphaericus* into *Bacillus thuringiensis* subsp. *israelensis* and their expression. *Appl. Environ. Microbiol.* **56**, 340–344 (1990).
- Park, H.-W., Hice, R. H. & Federici, B. A. Effect of promoters and plasmid copy number on Cyt1A synthesis and crystal assembly in *Bacillus thuringiensis*. *Curr. Microbiol.* **72**, 33–40 (2016).
- Wu, D. & Federici, B. A. A 20-kilodalton protein preserves cell viability and promotes CytA crystal formation during sporulation in *Bacillus thuringiensis*. *J. Bacteriol.* **175**, 5276–5280 (1993).
- Miyatake, H., Hasegawa, T. & Yamano, A. New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL). *Acta Crystallogr. D Biol. Crystallogr.* **62**, 280–289 (2006).
- Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 035108 (2012).
- Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. New Python-based methods for data processing. *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1274–1282 (2013).
- Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).

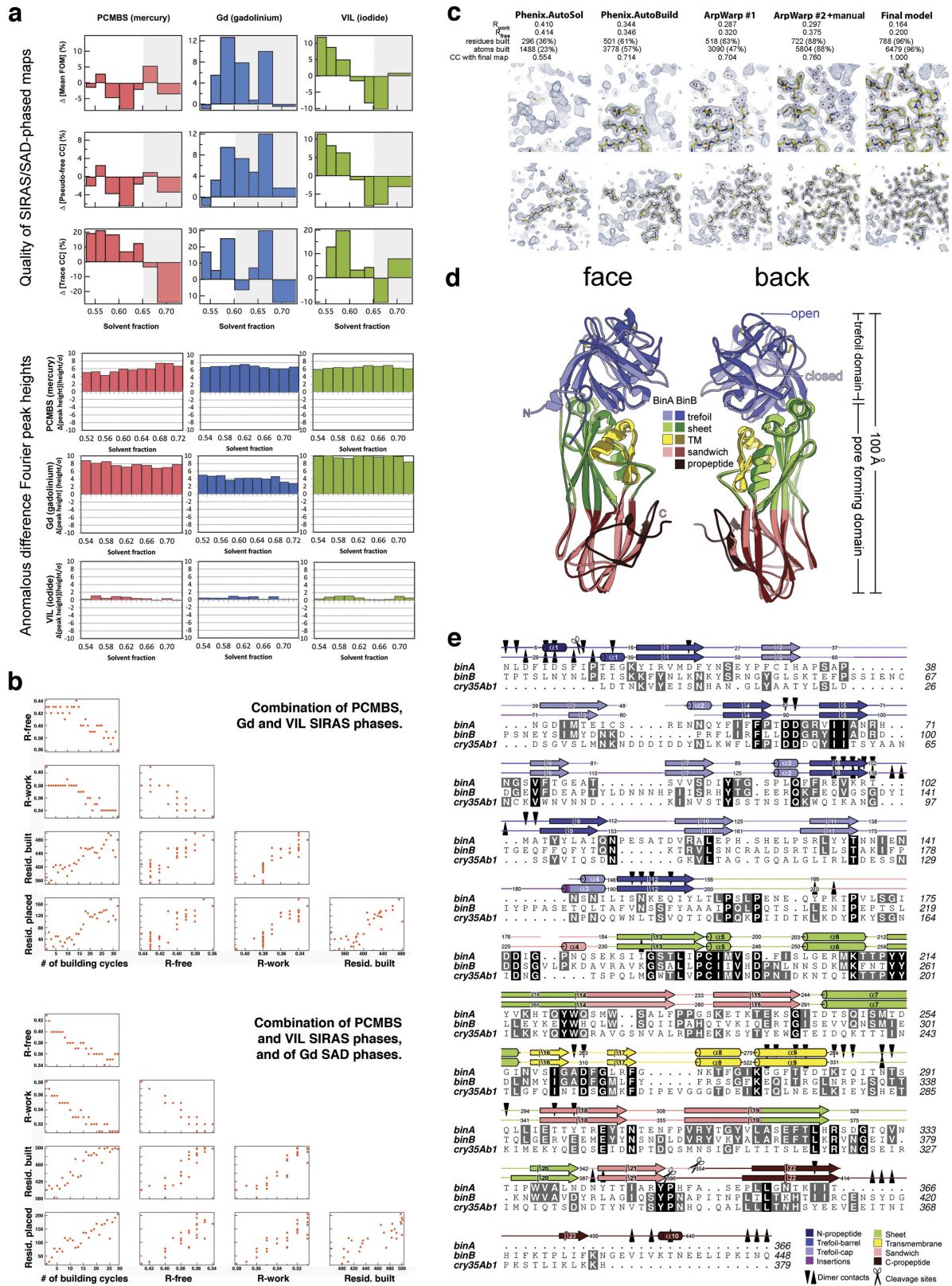
41. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
42. Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protocols* **3**, 1171–1179 (2008).
43. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
44. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374**, 22–37 (2003).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
47. Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2210–2221 (2004).
48. Ursby, T. & Bourgeois, D. Improved estimation of structure-factor difference amplitudes from poorly accurate data. *Acta Crystallogr. A* **53**, 564–575 (1997).
49. Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat. Protocols* **2**, 2728–2733 (2007).
50. Colletier, J.-P. *et al.* Shoot-and-Trap: use of specific x-ray damage to study structural protein dynamics by temperature-controlled cryo-crystallography. *Proc. Natl Acad. Sci. USA* **105**, 11742–11747 (2008).
51. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
52. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
53. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
54. Zeldin, O. B., Gerstel, M. & Garman, E. F. RADDose-3D: time- and space-resolved modelling of dose in macromolecular crystallography. *J. Appl. Cryst.* **46**, 1225–1230 (2013).
55. Cole, A. Absorption of 20-eV to 50,000-eV Electron Beams in Air and Plastic. *Radiat. Res.* **38**, 7–33 (1969).
56. Singkhamanan, K., Promdonkoy, B., Chaisri, U. & Boonserm, P. Identification of amino acids required for receptor binding and toxicity of the *Bacillus sphaericus* binary toxin. *FEMS Microbiol. Lett.* **303**, 84–91 (2010).
57. Kelker, M. S. *et al.* Structural and biophysical characterization of *Bacillus thuringiensis* insecticidal proteins Cry34Ab1 and Cry35Ab1. *PLoS One* **9**, e112555 (2014).
58. De Colibus, L. *et al.* Structures of lysenin reveal a shared evolutionary origin for pore-forming proteins and its mode of sphingomyelin recognition. *Structure* **20**, 1498–1507 (2012).
59. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
60. McCoy, A. J., Chandana Epa, V. & Colman, P. M. Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.* **268**, 570–584 (1997).



Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Data collection and heavy-atom substructure determinations. **a**, The ‘MESH-on-a-stick’ sample injector configuration (see Methods). In the three panels, the yellow X below the capillary indicates the X-ray path into the page. The middle panel shows a closer view of the injector tip; the right panel shows an on-axis view of the sample injecting during the experiment. **b**, Our choice of X-ray wavelength for diffraction and MIRAS phasing was a compromise between maximizing the heavy atom anomalous signals, f'' , as indicated by the curves for each element, and maximizing the number of data sets collected in the time allotted for the experiment. The grey bar corresponds to the wavelength we used, 1.41 Å. **c**, Difference Patterson maps calculated at 2.8 Å resolution. Sharpening (-5.0 Å^2) was applied to Hg and VIL maps. Coefficients for the PCMBs and VIL maps were obtained from both isomorphous and anomalous differences. The Gd difference Patterson map was calculated from anomalous differences only. Contours start at the 1.5σ level and continue at 0.5σ intervals. Peaks corresponding to vectors between heavy atoms stand out as high peaks, up to 7.5σ . **d**, Heavy-atom sites were located successfully for each of the three derivatives using the program SHELXD. We compared the quality of potential heavy-atom substructure solutions obtained from two sources of heavy-atom signal: single wavelength anomalous dispersion (SAD, red) and a combination of anomalous dispersion and isomorphous differences (SIRAS, blue). Ten-thousand independent trials were performed for each derivative and signal source. Each dot in the scatter plots indicates the quality of an

individual substructure solution. The vertical axis, labelled CC_{all} , indicates the consistency between the potential solution and the diffraction data as the correlation coefficient between normalized structure factors, E_{calc} and E_{obs} . The horizontal axis, labelled PATFOM (Patterson figure of merit), indicates the consistency between the observed difference Patterson map and that predicted by the potential solution. Successful substructure determination is suggested by the appearance of a sharp separation between two populations of potential solutions: a cluster with lower values of CC_{all} and PATFOM (incorrect solutions) and a cluster with higher values (correct solutions). Such is the case for all the trials performed, except for VIL using the SAD signal, where only a single population of solutions is observed. Evidently, the SAD signal was insufficient for accurate location of iodine sites. For VIL, we relied on the accuracy of sites obtained from the SIRAS signal. In most cases, the SIRAS (blue) signal is stronger than the SAD signal (red), indicating good isomorphism between native and derivative data sets. Only in the case of GdCl_3 does the SAD signal appear to be better than the SIRAS signal. The histograms in the right column indicate the number of potential substructure solutions with given values of CFOM (combined figure of merit). The histograms recapitulate the trends observed in the scatter plots. **e**, The correlation coefficient (CC_{iso}) measures the agreement and R_{iso} measures the discrepancy between the native structure factors and those of each of the derivatives. Each of our three derivative data sets shows isomorphism with the native data set up to 2.8 Å resolution.



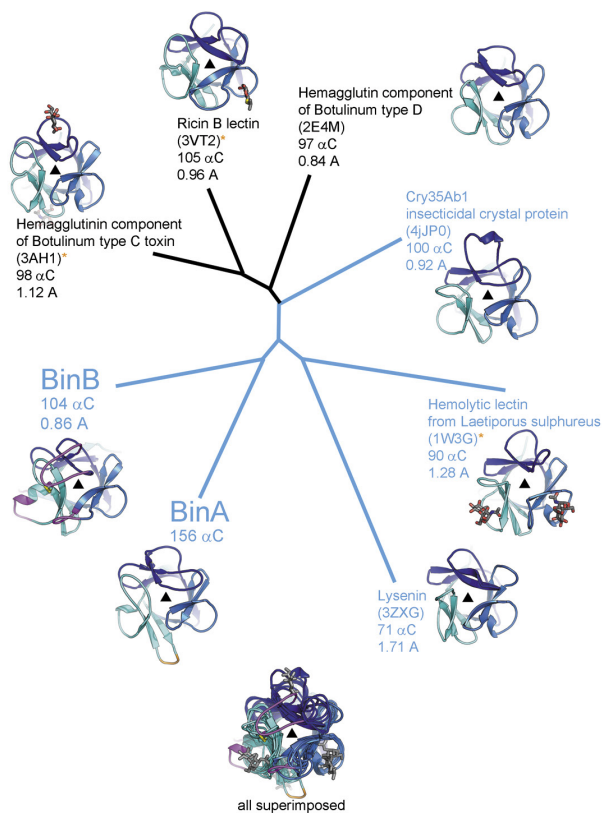
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Structure solution and model building.

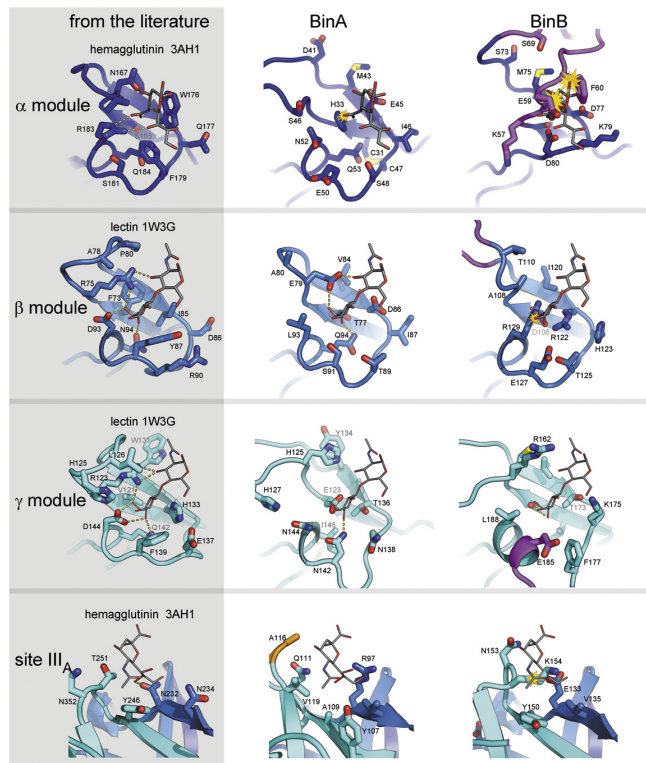
a, Evidence for choosing the correct hand of heavy atom substructures. We illustrate here the two types of comparison we used for choosing the correct hand of the heavy-atom substructures: (1) comparisons of the quality of the SIRAS and SAD phased maps (upper three panels) and (2) comparisons of the heights of anomalous difference Fourier peaks (lower three panels). These comparisons are made between maps calculated in opposite hands; the correct hand is indicated by the individual with higher positive values. The disparity in values (Δ) is indicated on the vertical axes of the graphs. Greater $|\Delta|$ values indicate a stronger phasing power and more reliable choice of hand. There are six comparisons shown for each of the three heavy derivatives: PCMBs (mercury) in red, GdCl_3 (gadolinium) in blue, and VIL (iodide) in green. The top three panels illustrate the percent difference between hands in the mean figure of merit (Mean-FOM), the pseudo-free correlation coefficient (Pseudo-free CC) of the density-modified map, and the correlation coefficient of the trace (Trace-CC) as reported by ShelxE. The sites and phases were obtained from SIRAS signal for mercury and iodide, and from SAD signal for Gd. The most probable solvent content is $\sim 59\%$, corresponding to one BinAB complex per asymmetric unit. However, we note conflicting choices of hand indicated by fluctuations in the sign of Δ accompanying small variations in the solvent content used in the density modification step (horizontal axes). We found that the difference Fourier maps (lower three panels) offered a stronger and more consistent indication of the choice of hand even when the statistics from SIRAS and SAD phased maps themselves differed little between hands. In these panels, SIRAS phases from the each heavy atom (3 columns) were used to compute three anomalous difference Fourier maps, using as coefficients, the anomalous differences from each derivative (rows 4–6). The value Δ corresponds to the height of the highest peak in the map computed in the original hand minus the corresponding peak in the map computed in the inverted hand. The graphs show that for all three derivatives, the original hand choice was correct (indicated by positive Δ), consistent across choices of solvent content (all Δ have the same sign within a graph), and consistent across sources of anomalous differences (all Δ have the same sign within a column). **b**, Automated tracing and model-building. Phases from the three derivatives were combined

using SOLVE⁴⁴ (see Methods). The hands which we decided on during the phasing step (**a**) were specified as ‘known’ to phenix.autosol. We then used RESOLVE⁴⁴ (see Methods) to trace the density and build a model. The upper and lower panels show the progress of model building, depending on whether anomalous and isomorphous differences were combined from all derivatives (upper panel) or mixed phases (that is, anomalous and isomorphous differences from PCMBs and VIL, and anomalous differences from the Gd derivative) (lower panel). Each panel shows scatter plots of R_{free} , R_{factor} , number of residues built and number of residues placed in sequence as a function of the number of cycles. The use of mixed phases allowed us to obtain a better model, faster (lower panel). **c**, Electron density maps at various stages of model building. From left to right, the panels illustrate progressive improvement in map and model at two representative regions of BinAB (upper versus lower panels). The number of residues built (including residues without side chains) is noted at each stage, as well as the number of protein atoms built. The quality of the maps at each stage is reported as a correlation coefficient with the map obtained from the final model. Approximately 60% of the total atoms in BinAB were built automatically. **d**, Comparison of BinA and BinB structures. Superposition of BinA (lighter colours) and BinB (darker colours) shows similarity between molecules, which superimpose with an r.m.s.d. of 1.7 \AA over 329 pairs of α -carbons. The ‘face’ view displays the surface involved in the BinA–BinB dimer interface, and the barrel subdomain of the trefoil is oriented towards the viewer. The back view displays the outward faces of the molecules, with the putative carbohydrate binding modules, in the cap subdomain, oriented towards the viewer. One of the largest structural differences is located in a surface loop on the back face, in the trefoil domain (blue). In BinB, a disulfide bond (Cys67–Cys161, yellow sticks) pins a surface loop (residues 60–74) away from the opening in the trefoil domain (open), whereas in BinA, the analogous loop (residues 34–46) is stabilized by a different disulfide bond (Cys31–Cys47, yellow sticks) to take a conformation that covers the opening in the trefoil domain (closed). **e**, Structure-based sequence alignment of BinA, BinB, and cry35Ab1. The secondary structures of BinA and BinB are shown above the sequences. Heterodimer contacts and cleavage sites are noted.

a

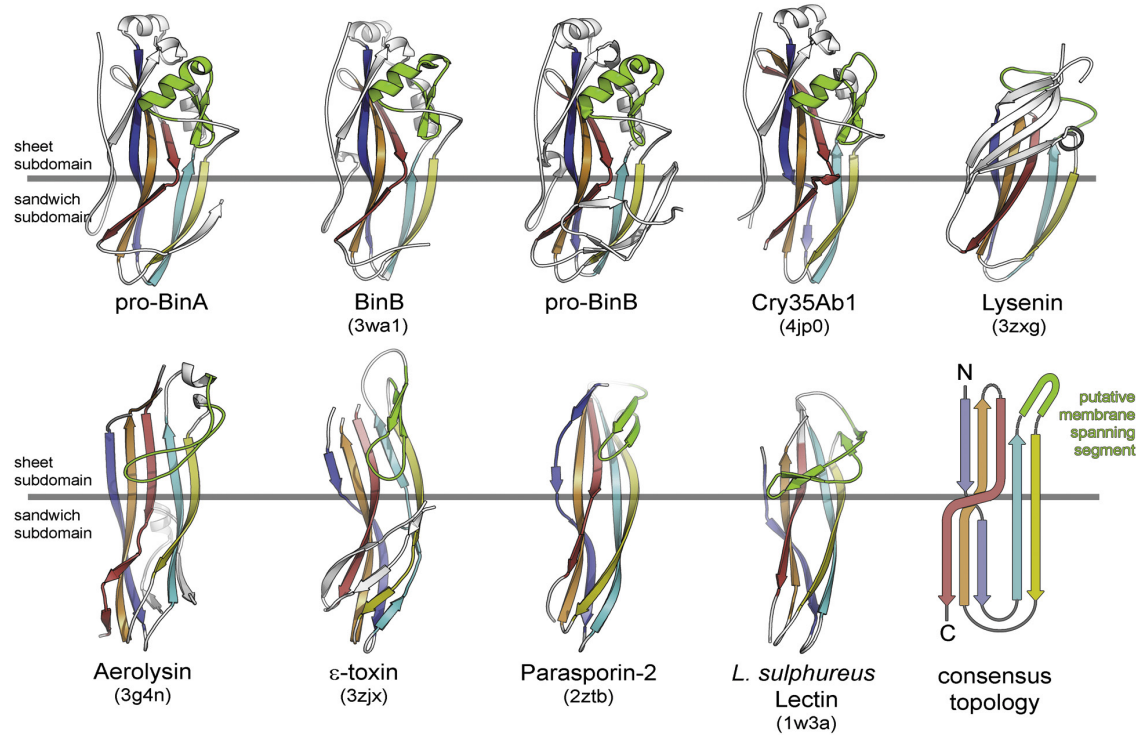
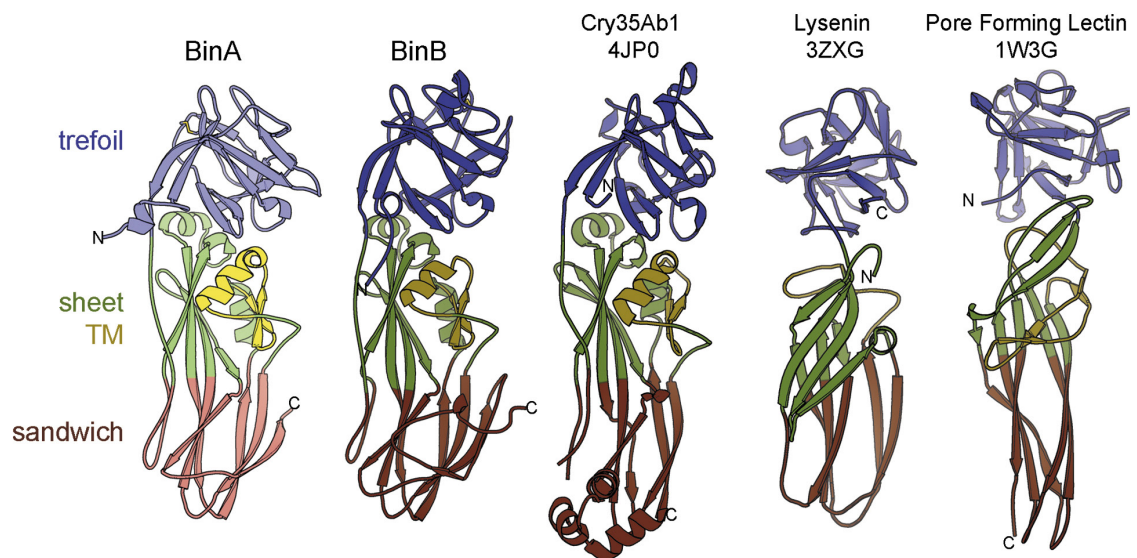
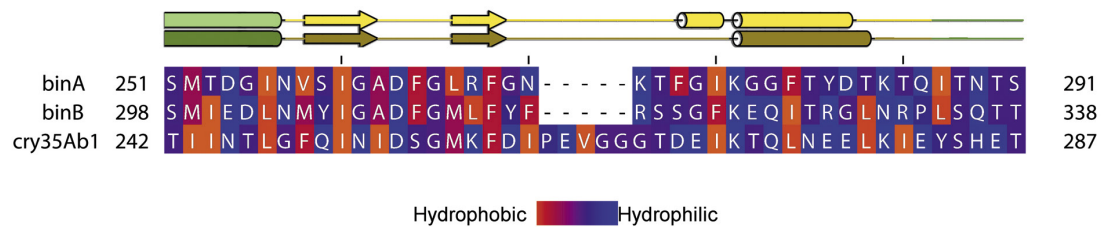


b



Extended Data Figure 3 | Trefoil domains of BinA and BinB.

a, Structural relationships among trefoil domains illustrated by a phylogenetic tree plot. Four of the structures used for comparison were identified from a structural similarity search through the Protein Data Bank conducted by the Dali server (using BinA residues 6–156 as the probe). The top four hits occupy the top half of the plot (3AH1, 3VT2, 2E4M, and 4JP0) and include the deadly toxin, ricin (3VT2). The remaining structures chosen for comparison (1W3G and 3ZXG) were selected based on their membership in the aerolysin family of toxins, of which BinA and BinB are members. That is, these are trefoils covalently linked to aerolysin-type pore-forming domains. These are highlighted in blue text and include another insecticidal protein from *B. thuringiensis*, Cry35Ab1 (4JP0). Note that BinA and BinB are nearly as distant from each other as they are from the closest homologues, haemagglutinin, ricin, and Cry35Ab1. Carbohydrate molecules are shown in sticks where coordinates are available. Notable loop insertions in BinA and BinB are coloured in orange and magenta, respectively. **b**, Carbohydrate-binding modules of BinA and BinB display different levels of structural integrity. No carbohydrates were included or observed in the crystals structure of BinAB. To investigate the structural integrity of the putative carbohydrate-binding pockets of BinAB, we superimposed coordinates of lectin (1W3G) and haemagglutinin (3AH1). The crystal structures illustrated in the left column are carbohydrate complexes chosen for their structural similarity to BinAB. Some modules appear competent for carbohydrate binding, such as the β - and γ -modules of BinA and the β -module of BinB. Others show steric clash (yellow starburst), such as the α -module of BinA and the β -module of BinB, which could be overcome by allowing adjustments in torsion angles. Notably, the α -module of BinB is completely occluded by the insertion in its sequence (magenta) and stapled shut by a disulphide bond. In addition to the canonical α -, β - and γ -binding modules, 3AH1 displays another weakly bound carbohydrate marked site IIIA (bottom panel). This site is illustrated here because its superimposed coordinates lie adjacent to Y150 in BinB. The Y150A mutation causes complete loss of receptor binding⁵⁶.

a**b****c**

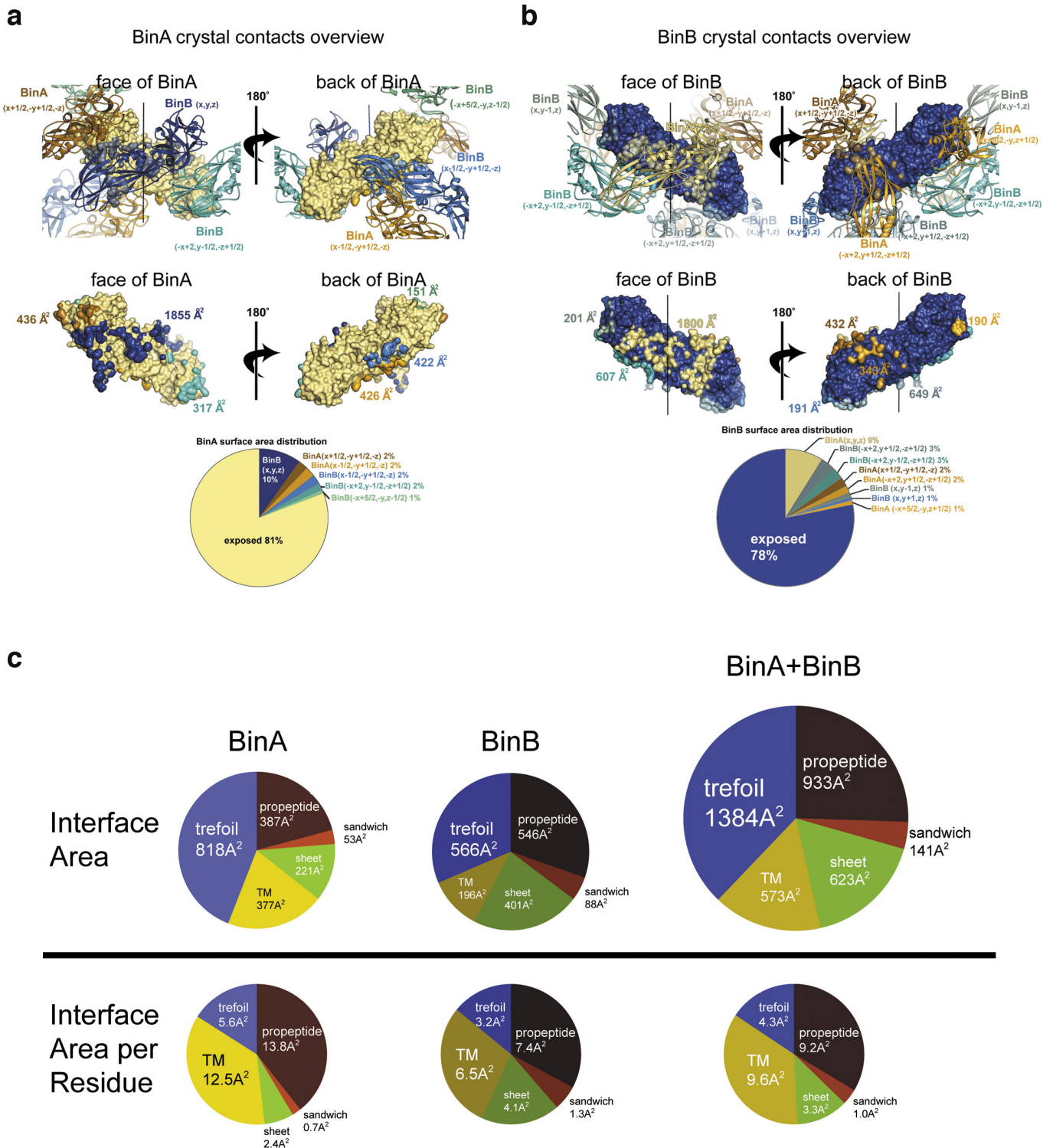
Extended Data Figure 4 | See next page for caption.

Extended Data Figure 4 | Pore-forming domains (PFD) of BinA**and BinB. a,** Topology of the aerolysin family of pore-forming toxins.

These share a core topology composed of five antiparallel β -strands and a putative membrane-spanning segment (green). PDB ID codes are included in parentheses. For clarity, we exclude from this illustration any accessory domains outside the pore-forming module (PFM) of these toxins. The PFM is divided into two subdomains: a β -sheet subdomain at one end (above the horizontal grey line) and a β -sandwich subdomain at the opposite end (below the horizontal grey line). The length, twist, and number of strands vary between toxins. Also, the putative membrane-spanning segment (green) varies widely in secondary structure. However, in all cases this putative membrane-spanning segment is located between the second and third strands, suggesting that these toxins might share a common mechanism of pore formation. **b,** Members of the aerolysin

family that also contain a β -trefoil domain like BinAB. These are:

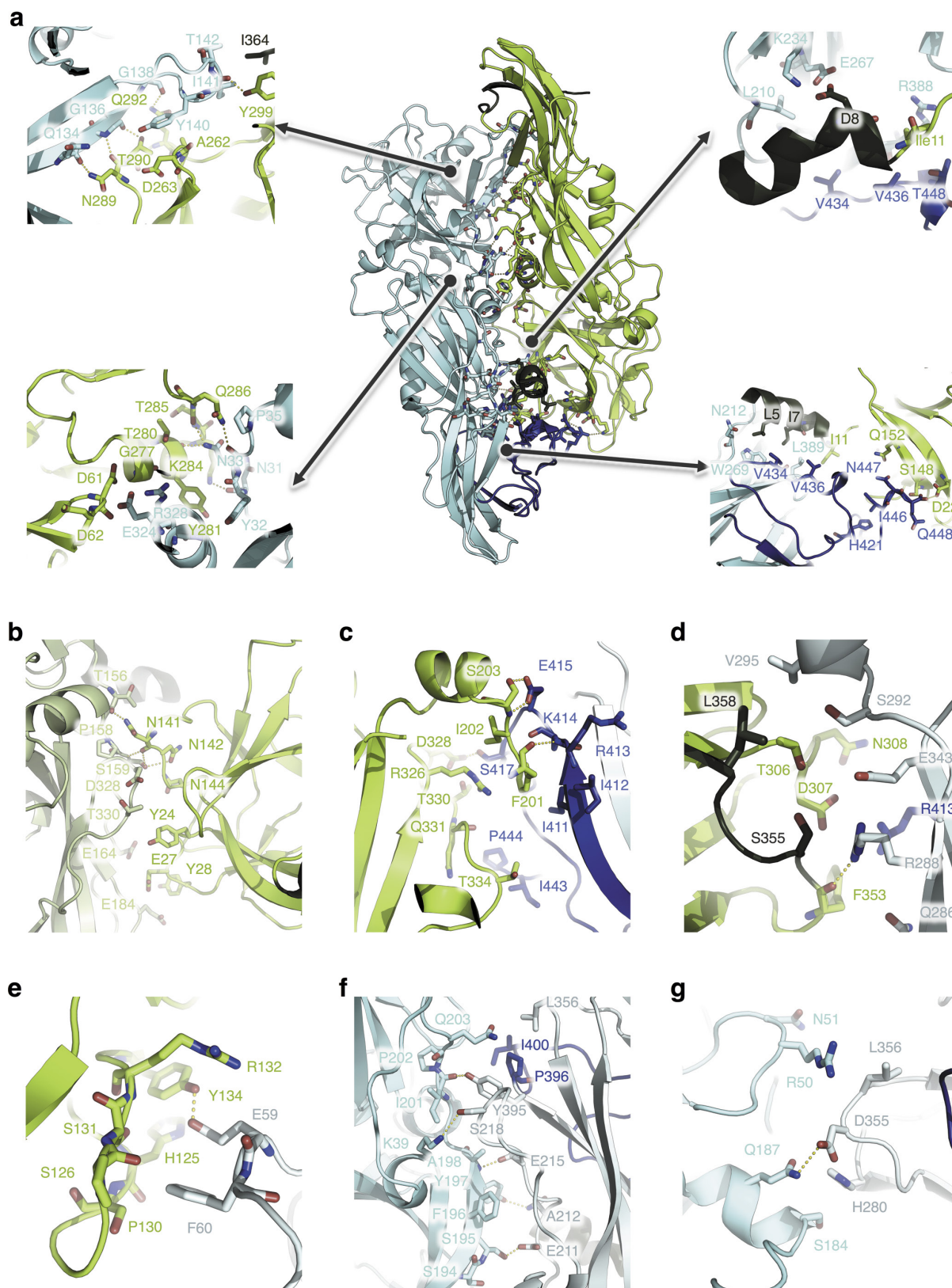
Cry35Ab1 toxin from *B. thuringiensis* (4jp0)⁵⁷, lysenin, a haemolytic toxin from the earthworm *Eisenia fetida* (3zxc)⁵⁸, and a pore-forming lectin from the mushroom *Laetiporus sulphureus* (1w3g)³⁰. **c,** Amphipathicity is evident in the sequence of the putative transmembrane (TM) subdomains of BinA and BinB. The observed secondary structures of BinA and BinB are shown above the sequence alignment. The range of the transmembrane subdomain is coloured yellow. Amino acids are coloured by hydrophobicity according to the scale given at the bottom. Note the alternating hydrophobic–hydrophilic pattern is especially prominent in the N-terminal half of the transmembrane subdomain. This pattern is consistent with the proposal of an oligomeric membrane-spanning β -barrel. The figure was made using the program Jalview⁵⁹.



Extended Data Figure 5 | See next page for caption.

Extended Data Figure 5 | Overview and analysis of molecular interfaces in the BinAB crystal. a, Overview of the six molecular interfaces involving BinA in the BinAB crystal. The reference copy of the BinA molecule is depicted as a beige molecular surface and its six neighbouring molecules are shown as cartoon ribbons (upper panels). Face and back views (left and right panels) reveal opposite surfaces of the BinA molecule. The largest interface is with BinB (x,y,z) which is shown most clearly in the face view (left panels) in dark blue. It is the only interface of the six that is large enough to stretch over most of the length of the molecule. In all views, the pseudo-two-fold axis relating BinA and BinB is in a vertical orientation (black line in upper panels). The areas of contact are illustrated on the BinA molecular surface (middle panels) in colours corresponding to the cartoon ribbons (upper panels). BinA molecules and surfaces are shown in beige shades; BinB molecules and surfaces are shown in blue-green shades. The pie chart shows the relative amount of total BinA surface area buried by each of the six crystal contacts and the remainder, which is solvent exposed. **b,** Overview of the eight molecular interfaces involving BinB in the BinAB crystal. The reference copy of the BinB molecule is depicted as a dark blue molecular surface and its eight neighbouring molecules are shown as cartoon ribbons (upper panels). Face and back views (left and right panels) reveal opposite surfaces of the BinB molecule. The largest interface is with BinA (x,y,z), shown most clearly in the face view (left panels) in beige. It is the only interface of the eight that is large

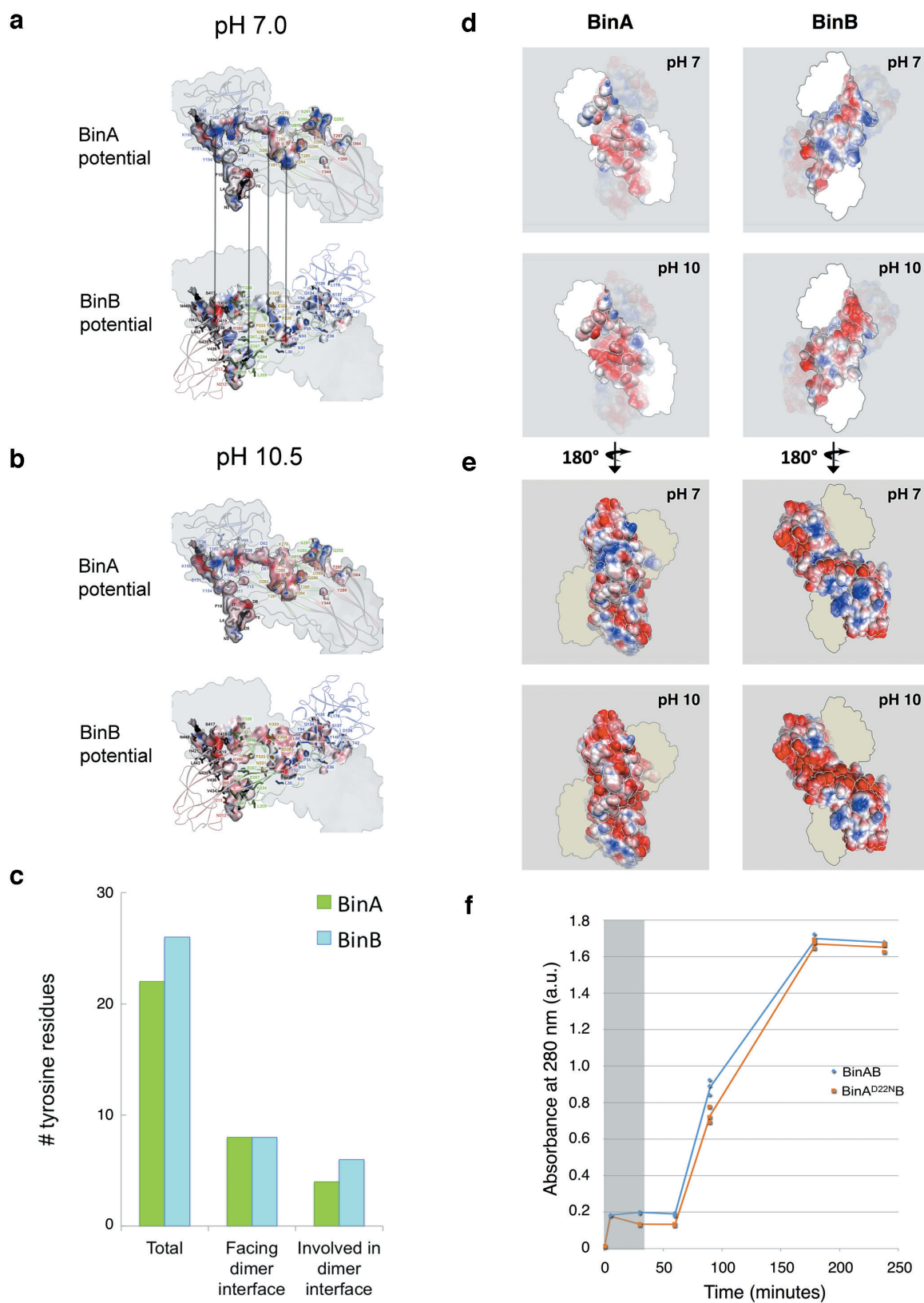
enough to stretch over most of the length of the molecule. In all views, the pseudo-two-fold axis relating BinA and BinB is in a vertical orientation (black line in upper panels). The areas of contact are illustrated on the BinA molecular surface (middle panels) in colours corresponding to the cartoon ribbons (upper panels). BinA molecules and surfaces are shown in amber shades; BinB molecules and surfaces are shown in blue-green shades. The pie chart shows the relative amounts of total BinB surface area buried by each of the eight crystal contacts and the remainder, which is solvent exposed. **c,** Distribution of the BinA–BinB interface area over its subdomains. The pie charts in the upper half show the area contributions to the principal BinA–BinB interface from each of the five named regions: trefoil domain, transmembrane subdomain, sheet subdomain, sandwich subdomain, and combined N- and C-terminal propeptides. The lower charts show analogous contributions on a per-residue basis. That is, the area contributed by each region is divided by the total number of residues comprising that region. These pie charts emphasize the role of the transmembrane subdomain in the dimer interface, perhaps to restrain this subdomain from inserting into a membrane until after the BinAB dimer dissociates. Notably, the higher efficiency of pore formation of BinA compared to BinB²⁰ correlates with the greater protection of its transmembrane domain (12.5 Å² buried per residue versus 6.5 Å² buried per residue) in the dimer.



Extended Data Figure 6 | See next page for caption.

Extended Data Figure 6 | Detailed views of the molecular interfaces in the BinAB crystal. a–g, BinA and BinB are shown as green and cyan ribbon diagrams, respectively. The C-terminal propeptide of BinB (residues 396–448) is highlighted in blue, while the N-terminal (residues 1–10) and C-terminal (residues 354–367) propeptides of BinA are shown in dark green. Contacting residues are shown as sticks. Polar interactions within a 3.6-Å cut-off are highlighted by yellow dashes. The contacts illustrated in panels **a–g** are detailed in Supplementary Tables 3–9, respectively. **a,** Molecular contacts between BinA (x, y, z) (green) and BinB (x, y, z) (cyan), that is, within the biological dimer. A large part of this interface involves the C-terminal propeptide of BinB. **b,** Molecular

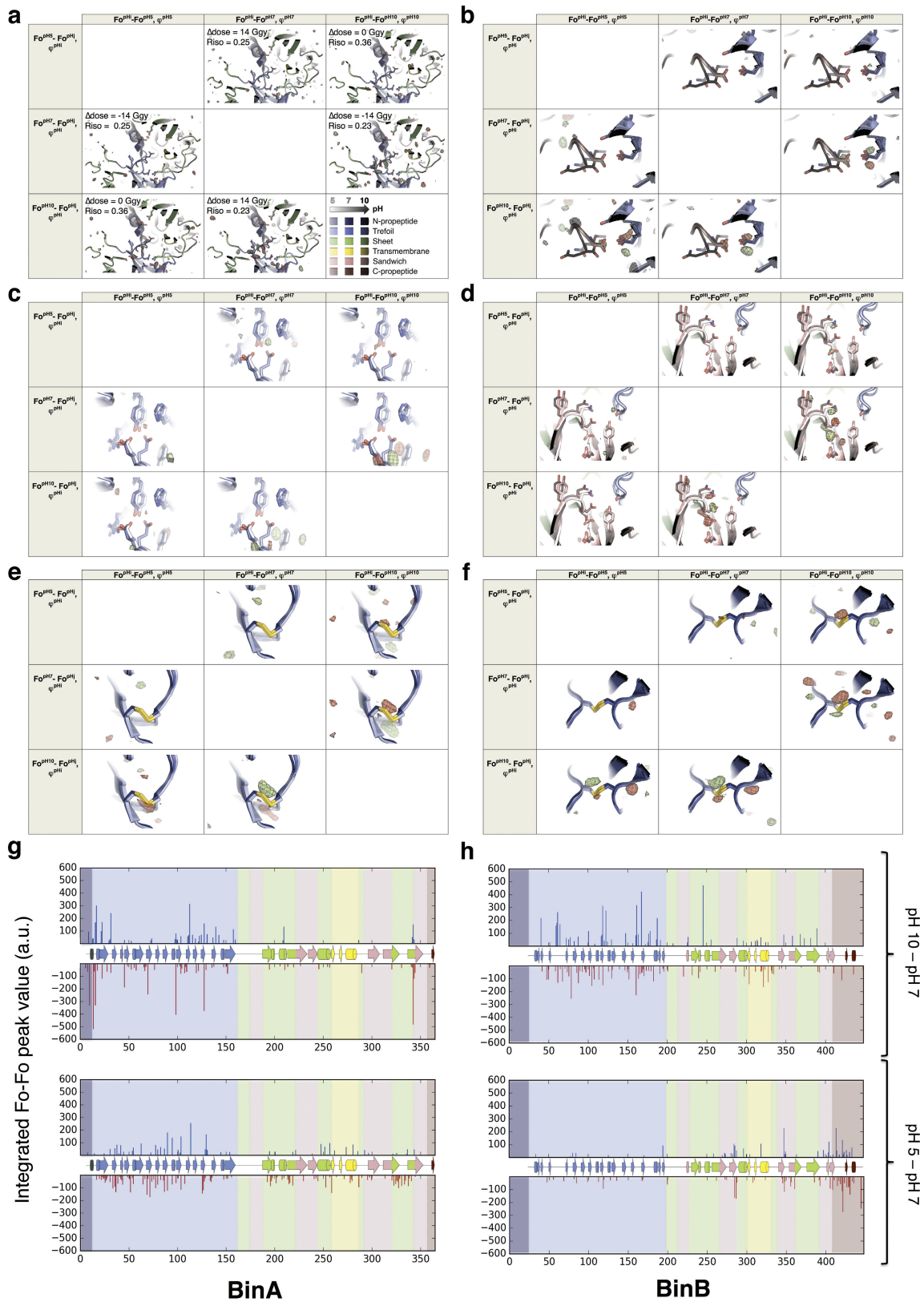
contacts between BinA (x, y, z) (green) and BinA ($x + 1/2, -y + 1/2, -z$) (lime green). **c,** Molecular contacts between BinA (x, y, z) (green) and BinB ($x - 1/2, -y + 1/2, -z$) (cyan). This interface involves the propeptide of BinB (residues 396–448). **d,** Molecular contacts between BinA (x, y, z) (green) and BinB ($-x + 2, y - 1/2, -z + 1/2$) (cyan). This interface involves the propeptide of BinB (residues 396–448). **e,** Molecular contacts between BinA (x, y, z) (green) and BinB ($-x + 5/2, -y, -z + 1/2$) (cyan). **f,** Molecular contacts between BinB (x, y, z) (cyan) and BinB ($-x + 2, y - 1/2, -z + 1/2$) (teal). A small part of this interface involves the propeptide of BinB (residues 396–448). **g,** Molecular contacts between BinB (x, y, z) (cyan) and BinB ($x, y - 1, z$) (teal).



Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Electrostatic complementarity, tyrosine distribution, predicted electrostatic changes upon pH elevation and crystal solubilisation assays. **a, b**, Electrostatic surface complementarity of the BinA–BinB interface. At pH 7.0 (**a**), complementary charges are notable between the BinA electrostatic surface potential (top) and the BinB electrostatic surface potential (bottom). The complementarity in potential is highlighted by the vertical arrows connecting adjacent patches on opposing surfaces of the interface. At pH 10.5 (**b**), deprotonation of tyrosine and increased negative charge on acid residues causes a reduction in electrostatic complementarity from 0.37 to 0.29 (ref. 60). All panels depict the BinA surface of the BinAB dimer interface. In the upper panels of **a** and **b**, this surface is coloured by electrostatic surface potential of BinA; in the lower panels, this surface is coloured by electrostatic surface potential of BinB. Residues lining the interface (sticks) are labelled with colour corresponding to the domain to which it belongs. The colour scheme is as described in Extended Data Fig. 2a. BinA residues are labelled in the upper panel. BinB residues are labelled in the lower panel. In all panels, the pseudo-two-fold axis relating BinA and BinB is in a vertical orientation (black line in Fig. 3a, lower panel). **c**, Distribution of tyrosine residues in the BinAB dimer. Of the total 49 tyrosine residues, 48 are ordered in the crystal structure. Of these, 20% are located in the dimer interface, which itself accounts for only 10% of the total molecular surface. Thus, the distribution of tyrosine residues is slightly more concentrated on the dimer interface compared to the remainder of the BinAB surface. Tyrosines outside the dimer interface are probably more prone to deprotonation than those within the dimer interface, due to differences in solvent accessibility. **d, e**, Electrostatic

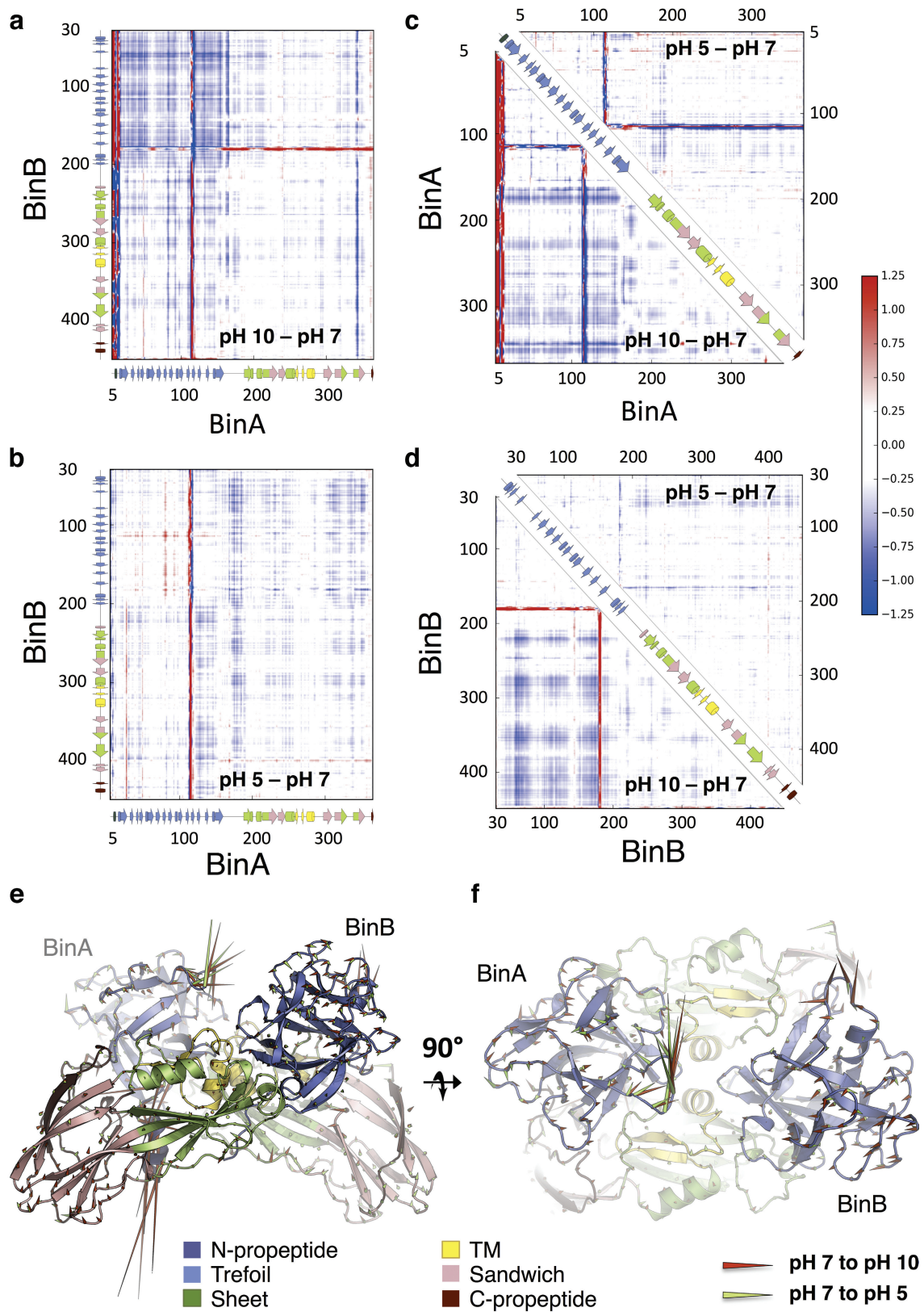
potential map of BinA and BinB. The surface of the BinAB dimer is depicted coloured by the electrostatic surface potential of BinA on the left, and by that of BinB on the right. In **d**, the regions of BinA (left) and BinB (right) that participate in the dimer interface are highlighted. In **e**, the external surface of the dimer is highlighted. In both **d** and **e**, the upper and lower panels show the electrostatic surface potentials of BinA (left) and BinB (right) at pH 7 and pH 10.5, respectively. **f**, Alkaline-induced crystal dissolution is delayed for the BinA D22N mutant compared to the wild type. Our structural data suggested that BinA Asp22 was an important pH sensor for triggering crystal dissolution at the high pH characteristic of the mosquito midgut. We reasoned that a D22N mutation in BinA would render the crystal less sensitive to pH by stabilizing a hydrogen bond with the BinB C-terminal carboxylate. We constructed a BinA D22N mutant and measured solubility of BinA D22N–BinB crystals at pH 10, collecting three data points for each time point. We found that its solubility *in vitro* decreased by 30% between 30 and 90 min at pH 10 compared to wild-type crystals, but not at pH 7 (individual measurements are plotted to indicate the range of variation). After 90 min, crystals of wild-type BinAB and BinA D22N–BinB are completely dissolved. This delay in crystal dissolution up to the 90-min time point is an important difference because in the fourth-instar *Culex* larvae, the larval feeding rate from the time particles are ingested until they are digested and exit the hindgut is 30 min (indicated by grey shading). Hence, the 60-min delay that we see in our experiments with D22N is long enough to contribute to the striking loss of toxicity of more than 20-fold at the LC95 level (Supplementary Table 13). These results are consistent with the model of Asp22 serving as a pH sensor for crystallization.



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Comparison of $F_o^{pH5}-F_o^{pH7}$ maps obtained from crystals receiving different X-ray doses suggests that the structural changes observed are due to pH change and not radiation damage. **a–f**, The pH 5 and pH 10 data sets were collected with a ~ 500 -fold higher dose than the pH 7 data set, raising the concern that some of the peaks in the $F_o^{pH10}-F_o^{pH7}$, φ^{pH7} map result from radiation damage, most notably those observed on disulfides. Panels **a–d** show, for each of the four regions identified as highly sensitive to pH elevation in Fig. 4c–f, the six possible $F_o^{pHj}-F_o^{pHi}$, φ^{pHi} maps calculated from the pH 5, pH 7 and pH 10 data sets and structures. Panels **e** and **f** show these maps around the disulfides of BinA (**e**) and BinB (**f**). BinA and BinB are shown as cartoons, coloured by subdomain, as in Fig. 4. The cartoons range in colour from pale to medium to dark, signifying the pH values 5, 7, and 10, respectively. Consistent with the hypothesis that a 500-fold difference in dose causes no major structural change, we see a lack of peaks in the $F_o^{pH5}-F_o^{pH7}$, φ^{pH7} map ($R_{iso}=0.26$) around disulfides (**e**, **f**) and other pH-sensitive residues (**a–d**). Consistent with the hypothesis that the peaks observed in the $F_o^{pH10}-F_o^{pH7}$, φ^{pH7} map ($R_{iso}=0.23$) are caused by pH change, these peaks are reproduced in the $F_o^{pH10}-F_o^{pH5}$, φ^{pH5} and $F_o^{pH5}-F_o^{pH10}$, φ^{pH10} maps ($R_{iso}=0.35$). We interpret this pattern of peaks

as implying movement of the disulfide bonds rather than their disruption. This movement accompanies pH-sensitive rigid-body motion of the trefoil domains (Extended Data Fig. 9). **g**, **h**, Peaks stronger than $\pm 3.5\sigma$ were integrated contiguously in the $F_o^{pH10}-F_o^{pH7}$, φ^{pH7} (upper panels) and $F_o^{pH5}-F_o^{pH7}$, φ^{pH7} maps (lower panels) and then assigned to the closest residue. The secondary structures of BinA (**g**) and BinB (**h**) are shown as cartoons, coloured by subdomain as in Fig. 4. The background of the sequence is also coloured by subdomain. The sequence-wise integration of the $F_o^{pH10}-F_o^{pH7}$, φ^{pH7} map reveals that BinB is more affected by the pH elevation than BinA, and in both chains, the trefoil is more affected than the pore-forming domain. The propeptide and transmembrane regions of both proteins are also sensitive to pH elevation. Peaks in the $F_o^{pH5}-F_o^{pH7}$, φ^{pH7} map (lower panels) are smaller in magnitude and concentrated in the trefoil domain of BinA and the C-terminal propeptide of BinB. They correspond to side-chain reorientation rather than increased dynamics or domain motion (Extended Data Fig. 9). The marked difference in pattern between the $F_o^{pH10}-F_o^{pH7}$, φ^{pH7} and $F_o^{pH5}-F_o^{pH7}$, φ^{pH7} map integrations is consistent with the hypothesis that the peaks observed in these maps are not due to radiation damage, but rather to pH-induced conformational changes.



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Conformational changes in the BinAB dimer upon pH elevation from 7 to 10. a–d, Distance difference matrices (DDMs) calculated between the pH 7 (reference) structure and either the pH 10 or the pH 5 structure. Blue and red indicate decreases and increases in C α –C α distances in the pH 10 or pH 5 structures as compared to the pH 7 structure, respectively. The secondary structures of BinA (**a–c**) and BinB (**a, b, d**) are recapitulated by cartoons on the side or the diagonal of the DDMs. These cartoons are coloured by subdomain as in Fig. 4. **a,** Intermolecular (BinA versus BinB) DDM between the pH 10 and the pH 7 structures. This DDM illustrates that the BinAB dimer contracts upon pH elevation, with the two trefoil domains coming closer to one another. This might be due to electrostatic repulsion at crystal contact zone 5 (Fig. 4e, Extended Data Fig. 6e and Supplementary Table 7), which involves the trefoils of BinA and BinB from two symmetry-related dimers. **b,** Intermolecular (BinA versus BinB) DDM between the pH 5 and the pH 7 structures. The pH 5 structure is overall slightly more compact than the pH 7 structure but shows no major conformational changes. **c, d,** Intramolecular DDMs of BinA (**c**) and BinB (**d**). Changes in C α –C α distances between the pH 10 and the pH 7 structures are reported below the diagonal, while those between the pH 7 and the pH 5 structures are shown above the diagonal. The pH 5 and pH 7 structures of BinA

(**c**) and BinB (**d**) are overall similar, with only the BinA loop Ile110–Arg120 and BinB loop Lys175–Ser184 showing a noticeable difference in conformation. In contrast, the pH 10 structures of BinA (**c**) and BinB (**d**) appear more compact. On the local level, striking conformational changes are observed upon pH elevation in the N-terminal propeptide of BinA, in loops Ile110–Thr120 (trefoil) and Asn341–Tyr345 (PFD) of BinA, and in loop Lys175–Ser184 (trefoil) of BinB. The increase in compactness is due to the trefoil domain coming closer to the PFD in both BinA and BinB. BinA loop Ile110–Thr120 appears sensitive to both increases and decreases in pH. **e, f,** Porcupine plots depicting differences between structures of BinAB for pH 7 versus pH 5 (green arrows) and pH 7 versus pH 10 (red arrows). The pH 7 structure of BinAB is shown, coloured by subdomain as in Fig. 4. The movement of C α atoms is indicated by arrows on the ribbon representation, with the magnitude of motions illustrated by length of arrows exaggerated by 2.5 Å to increase visibility (for all atoms that move by more than 0.1 Å). **e,** View of the BinAB dimer, in an orientation similar to Fig. 4b. As compared to Fig. 3a, b, this view is rotated by 180° around the vertical axis. **f,** View from the top of the trefoil domains; this face of the BinAB dimer is presumably that interacting with the apical membrane of larvae midgut cells. The view in **f** is 90° apart from that in **e**.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	Native, pH 7 5FOY ^a	PCMBs ^b	Gd ^c	VIL ^d	Native, pH 10 5FOZ ^e	Native, pH 5 5G37 ^f
Data collection						
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	86.9, 97.4, 128.3	87.1, 97.8, 128.4	87.0, 97.5, 128.1	86.7, 97.3, 127.7	86.7, 97.3, 127.7	86.8, 97.0, 127.0
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Wavelength (Å)	1.212	1.408	1.408	1.408	1.459	1.459
X-ray beam focus (μm)	1.3	1.3	1.3	1.3	0.25	0.25
Photons/pulse (x 10 ¹¹)	1.9	7.2	7.2	7.2	6.4	6.4
Pulse duration (fs)	31.9	45.1	45.1	45.1	40.8	40.8
Absorbed dose (GGy) ^g	0.03	0.15	0.15	0.15	14.4	14.4
Number of collected frames	312,659	761,052	431,986	264,770	371,386	493,979
Number of indexed patterns	41,206	184,091	131,137	42,537	27,792	17099
Number of indexed images accepted by <i>cctbx.prime</i>	40,794	170,616	123,484	39,789	26,022	16224
Resolution (Å) ^h	43.5-2.25 (2.29-2.25)	43.5-2.40 (2.44-2.40)	43.5-2.35 (2.39-2.35)	43.5-2.60 (2.64-2.60)	43.5-2.37 (2.44-2.37)	43.5-2.50 (2.54-2.50)
Number of observations ⁱ	12,854,588	53,361,546	39,314,393	11,249,082	12,069,895	5,495,157
I/σI	2.4 (0.5)	2.8 (0.7)	3.1 (0.6)	3.7 (0.9)	2.7 (0.9)	6.34 (2.46)
CC _{1/2}	96.8 (26.2)	98.6 (16.2)	98.4 (3.5)	96.3 (17.5)	97.3 (34.1)	86.8 (28.0)
Completeness (%)	99.7 (99.7)	99.8 (96.9)	99.8 (97.0)	99.4 (96.2)	99.9 (100.0)	99.9 (89.4)
Multiplicity	65.9 (4.8)	372 (4.3)	260.1 (4.4)	100.4 (4.7)	88.8 (19.2)	76.62 (8.61)
Refinement						
Refinement target function	Phased maximum- likelihood				Maximum- likelihood	Maximum- likelihood
Resolution (Å)	2.25 (2.30 – 2.25)				2.40 (2.46 – 2.40)	2.50 (2.56 – 2.50)
Number of reflections	52379 (3523)				42817 (2923)	37785 (2557)
R _{work} / R _{free} ^g	0.164 (0.288) / 0.200 (0.327)				0.165 (0.264) / 0.211 (0.311)	0.211 (0.357) / 0.262 (0.427)
Number of atoms						
Protein	6479				6350	6415
Water	571				730	720
B-factors (Å ²)						
Protein (BinA/BinB)	47.5 / 40.7				52.8 / 47.3	39.2 / 36.8
Water	50.8				62.4	41.01
R.m.s. deviations						
Bond lengths (Å)	0.006				0.007	0.002
Bond angles (°)	1.359				0.8	0.5

^a41,206 crystals were used to produce the native, pH 7 data set.^b184,091 crystals were used to produce the PCMBs data set.^c131,137 crystals were used to produce the Gd data set.^d42,537 crystals were used to produce the VIL data set.^e27,792 crystals were used to produce the native, pH 10 data set.^f17,099 crystals were used to produce the native, pH 5 data set.^gAs calculated by Raddose-3D⁵⁴. Note that these dose calculations do not yet take into account the escape of photoelectrons from the diffracting volume. The track length in protein and water of photoelectrons has been estimated to 3 μm, for 1 Å wavelength X-rays⁵⁵, suggesting that most photoelectrons travel outside of the diffracting volume without losing all their energy within it (crystals are 0.25 × 0.35 × 0.75 μm³ on average). Although these are not the only electrons to contribute to the damage, a preliminary photoelectron escape model incorporated into an unreleased version of RADDose-3D suggests that the absorbed dose is overestimated by a factor of ~100 (J. Brooks-Bartlett and E. Garman, personal communication) for these small crystals.^hValues in parentheses are for highest-resolution shell.ⁱIncluding negative intensity observations.

Arginine phosphorylation marks proteins for degradation by a Clp protease

Débora Broch Trentini^{1*}, Marcin Józef Suskiewicz^{1*}, Alexander Heuck¹, Robert Kurzbauer¹, Luiza Deszcz¹, Karl Mechtler^{1,2} & Tim Clausen¹

Protein turnover is a tightly controlled process that is crucial for the removal of aberrant polypeptides and for cellular signalling. Whereas ubiquitin marks eukaryotic proteins for proteasomal degradation, a general tagging system for the equivalent bacterial Clp proteases is not known. Here we describe the targeting mechanism of the ClpC–ClpP proteolytic complex from *Bacillus subtilis*. Quantitative affinity proteomics using a ClpP–trapping mutant show that proteins phosphorylated on arginine residues are selectively targeted to ClpC–ClpP. *In vitro* reconstitution experiments demonstrate that arginine phosphorylation by the McsB kinase is required and sufficient for the degradation of substrate proteins. The docking site for phosphoarginine is located in the amino-terminal domain of the ClpC ATPase, as resolved at high resolution in a co-crystal structure. Together, our data demonstrate that phosphoarginine functions as a bona fide degradation tag for the ClpC–ClpP protease. This system, which is widely distributed across Gram-positive bacteria, is functionally analogous to the eukaryotic ubiquitin–proteasome system.

Proteins destined for degradation are removed by energy-dependent proteases such as the eukaryotic proteasome or the bacterial Clp proteolytic complexes¹. In these tightly regulated protein shredders, the proteolytic sites are sequestered within an inner chamber that is only accessible through axial entrance gates². The gates are in turn guarded by regulatory AAA (ATPases associated with diverse cellular activities) complexes that are responsible for recognizing substrate proteins as well as for unfolding and translocating them into the protease cage^{3,4}. In eukaryotes, substrate selection depends on a range of ubiquitin ligases that mark substrates with a polyubiquitin tag, a degradation signal recognized by the AAA regulatory particle of the 26S proteasome^{5,6}. An analogous system involving the protein modifier Pup targets substrates to the eukaryotic-like core proteasome present in mycobacteria and closely related species^{7,8}. However, it is not known whether the ATP-dependent Clp proteases, which are found in almost all bacteria, require a general post-translational tagging system. The Clp complexes are thought to use the N-terminal domains (NTD) of the AAA ATPases (ClpA, ClpC, ClpE or ClpX) to recognize specific degradation motifs, known as degrons, which are typically located at the N- or C-terminal ends of target proteins^{9–11}. These degrons can also be introduced by the specialized SsrA tagging system, which is used for rescuing stalled ribosomes¹². Alternatively, substrate recruitment may be aided by adaptor proteins that tether selected substrate proteins to the Clp proteolytic complex, thus facilitating their degradation⁹.

In *B. subtilis* and other Gram-positive bacteria, the ClpC–ClpP (ClpCP) protease, which is constituted by the AAA unfoldase ClpC and the protease ClpP, is an important proteolytic machine for eliminating unfolded and aggregated proteins. The ClpCP proteolytic complex is under the control of McsB, the founding member of a class of protein kinases targeting arginine residues¹³. First, McsB controls the amounts of ClpCP in the cell by phosphorylating and inhibiting the transcriptional repressor CtsR^{13–15}, which in turn regulates *clpC* and *clpP* gene expression¹⁶. Second, McsB has been reported to function as an adaptor protein of ClpC by stimulating its ATPase activity and promoting degradation of the CstR substrate^{17,18}. In addition to regulating CtsR

and ClpC, McsB phosphorylates hundreds of diverse proteins *in vivo*, as revealed by *B. subtilis* phosphoproteomic analyses^{15,19}. This promiscuous activity suggested a more general function of the protein arginine kinase in the stress response of Gram-positive bacteria.

ClpCP degrades pArg proteins *in vivo*

Arginine residues are frequently observed at molecular interfaces crucial for protein folding and assembly²⁰. Therefore, arginine phosphorylation, resulting in a net-charge inversion, is predicted to have a strong effect on protein stability. Of note, the kinase catalysing this reaction, McsB, has many substrates *in vivo* and is transcriptionally co-regulated with ClpP, the major protease of *B. subtilis*. We thus proposed that arginine phosphorylation may have a direct role in the degradation of aberrant proteins. To test this assumption, we monitored the fate of phosphoarginine (pArg) proteins *in vivo* by expressing an inactive trapping variant of the ClpP protease (Ser98Ala, ClpP^{TRAP}; refs 10, 21). Substrates captured within the protease cage can be co-purified and analysed by mass spectrometry (MS). To perform the pull-down experiments in the wild-type *B. subtilis* background, we engineered a ClpP mutant that does not interact with the endogenous, active protease. For this purpose, we exchanged residues of an ion pair at the interface of the ClpP heptamer. The resulting cross mutant (Glu119Arg/Arg142Glu, ClpP^X) did not form heterooligomers with wild-type ClpP, but maintained the ability to assemble a substrate-trapping cage (Extended Data Fig. 1).

Quantitative MS analysis of ClpP pull-downs from heat-shocked bacteria (Extended Data Fig. 2a, b, strategy illustrated in Fig. 1a) revealed a large number of proteins that were specifically captured by the ClpP^{X-TRAP} mutant (Fig. 1b and Supplementary Table 1). Despite the technical difficulties in identifying arginine phosphorylations¹⁵, we detected 13 pArg proteins among the 233 isolated ClpP substrates (Extended Data Table 1, Supplementary Table 2). Taking into account the functional connection between McsB and ClpC, which are found in the same operon, we next asked whether pArg proteins are transferred into the ClpP cage by ClpC. To this end, we performed ClpP^{X-TRAP} pull-down analyses from

¹Research Institute of Molecular Pathology (IMP), Dr-Bohr-Gasse 7, 1030 Vienna, Austria. ²Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Dr-Bohr-Gasse 3, 1030 Vienna, Austria.

*These authors contributed equally to this work.

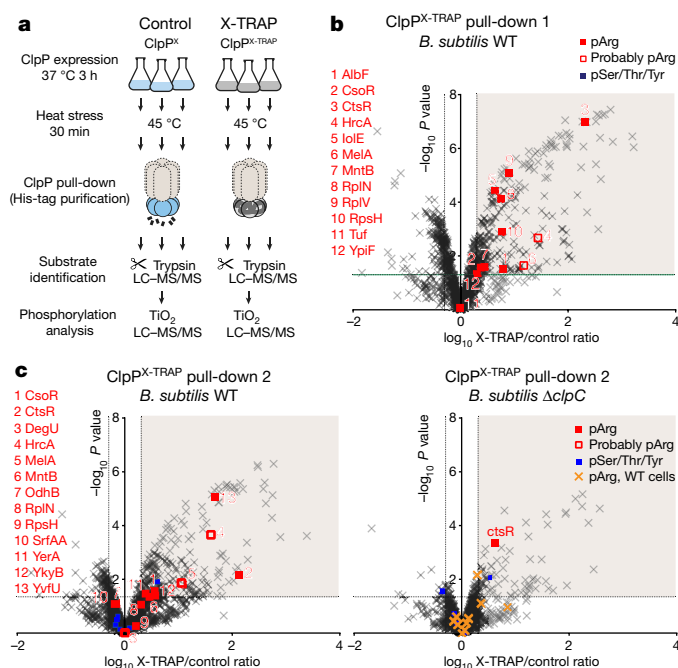


Figure 1 | Pull-down of ClpP trapping mutants. **a**, Cartoon illustrating the experimental workflow. As indicated, all pull-down experiments were done in triplicates. LC-MS/MS, liquid chromatography tandem mass spectrometry. **b**, Volcano plot illustrating proteins identified in ClpP^X (control) and ClpP^{X-TRAP} pull-downs after expression in a *B. subtilis* wild-type (WT) strain. Proteins were considered as ClpP substrates (shaded area) when the X-TRAP/control relative protein intensity (x axis) was >2 and the corresponding limma *P* value (y axis) was <0.05. Phosphorylated proteins are shown as filled squares (red: pArg, blue: pSer/Thr/Tyr). In a few cases, the phosphorylated residue could not be unambiguously localized. As the same phosphopeptides have been observed to contain a pArg in previous experiments, they are labelled as probably pArg (open red squares). Identified pArg proteins are listed on the left. **c**, Volcano plots of the ClpP pull-downs performed in *B. subtilis* wild-type and $\Delta clpC$ strains in parallel. For comparison, pArg proteins identified in the *B. subtilis* wild-type pull-downs are marked in orange in the $\Delta clpC$ plot.

wild-type and *clpC* knockout ($\Delta clpC$) cells in parallel. Whereas we observed 14 pArg proteins in the pull-downs performed in wild-type cells, the pull-downs in $\Delta clpC$ cells revealed only a single pArg substrate (Fig. 1c, Extended Data Table 1). Of note, this substrate, CtsR, was also shown to be targeted to ClpP by the ClpX and ClpE unfoldases^{22,23}. The almost complete absence of ClpP-trapped pArg proteins in $\Delta clpC$ bacteria is even more remarkable, as the deletion of ClpC increases the overall amounts of pArg proteins. Despite the presence of YwlE, a highly active arginine phosphatase preventing pArg identification in *B. subtilis* wild-type cells^{15,19,24}, phosphoproteomics analysis of $\Delta clpC$ cell lysates revealed 25 pArg sites (Supplementary Table 3). This finding highlights the active role of ClpC in directing pArg proteins to ClpP-dependent proteolysis. Consistent with the proposed model, $\Delta clpP$ *B. subtilis* cell extracts also accumulated pArg proteins (Supplementary Table 4). To estimate the fraction of pArg proteins among the ClpP substrates, we analysed the overlap of the ClpP degradome (as defined by our pull-down experiments) and the pArg proteome (sites detected previously^{15,19,25} and in the $\Delta clpP$ mutant strain). Accordingly, 25% of the proteins degraded by ClpP are substrates of MscB and thus potential candidates of the pArg-dependent degradation pathway (see also Supplementary Discussion and Extended Data Fig. 2c).

Protein phosphorylation stimulates ClpCP

To analyse how ClpC selects pArg-containing substrates, we reconstituted the ClpCP–McsB system *in vitro*. MscB was previously described as an adaptor of ClpC targeting the transcriptional repressor CtsR for

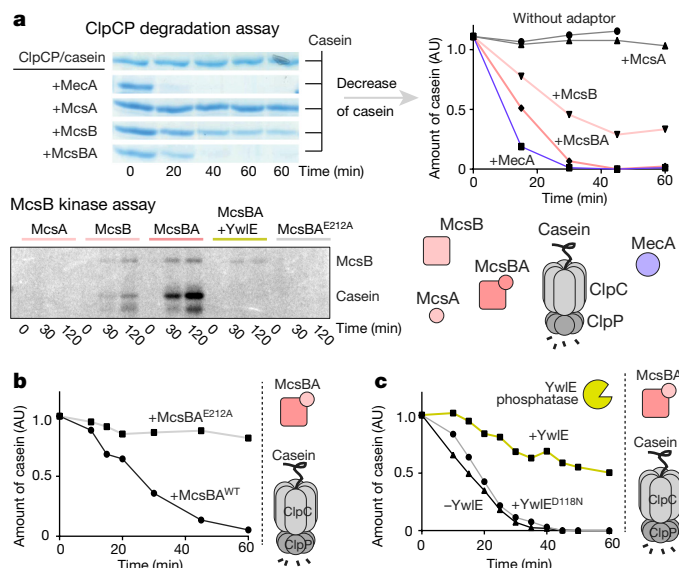


Figure 2 | Effect of McsB on the activity of ClpCP *in vitro*. **a**, ClpCP-mediated degradation of β -casein in the presence of different effector proteins. Here and in the following, a quantification of the β -casein band is presented (original SDS-PAGE gels in Supplementary Fig. 2). Bottom panel shows the kinase activity of assayed McsB variants in an autoradiography plot. AU, arbitrary units. **b**, In contrast to active McsB, the inactive McsB^{E212A} kinase cannot stimulate casein degradation by ClpCP. **c**, Effect of the YwlE arginine phosphatase on ClpCP activation by McsBA. The inactive YwlE^{D118N} mutant was used as a negative control.

ClpCP-mediated proteolysis^{17,18}. Although the kinase activity of MscB was shown to be required for CtsR degradation, it was not clear whether MscB itself, the substrate or the ClpCP protease became phosphorylated, and how this phosphorylation event enhanced protease activity. We thus recapitulated the corresponding kinase and protease assays using the intrinsically unfolded protein β -casein as a model substrate. We observed that the ClpCP protease complex alone was not active. However, in the presence of Meca, a well-characterized ClpC adaptor²⁶, the substrate was efficiently hydrolysed (Fig. 2a). Similarly, MscB induced the degradation of β -casein by ClpCP. The stimulatory effect was enhanced by the MscB activator MscA²⁷, which showed no effect on casein degradation by itself (Fig. 2a). To test whether the kinase activity of MscB is required for β -casein degradation, we used an inactive mutant of MscB (Glu212Ala; ref. 13), and, in parallel, probed the effect of the YwlE arginine phosphatase. Both kinase inactivation and phosphatase addition prevented substrate degradation (Fig. 2b, c), highlighting the importance of the arginine kinase activity of MscB for activating the ClpCP protease. This functional coupling is also reflected in the different kinase activities of the tested MscB variants (Fig. 2a).

pArg is a degradation tag for ClpCP

To explore the stimulatory effect of MscB further, we performed degradation assays in the presence of the free amino acid phosphoarginine (pArg^{AA}, in which 'AA' denotes the amino acid). We reasoned that pArg^{AA} may compete with, and thus reveal, the pArg-dependent activation event. When incubated with MscB and ClpCP, pArg^{AA} reduced the rate of β -casein degradation (Fig. 3a). We next explored the influence of pArg^{AA} on β -casein degradation by the ClpCP–Meca complex that should operate in a phosphorylation-independent manner. Unexpectedly, however, pArg^{AA} also inhibited (and to a greater extent) the activity of the Meca-stimulated ClpCP protease (Fig. 3b). By contrast, unphosphorylated arginine or phosphate did not block degradation. As it is known that effector proteins such as Meca dock to the NTD of AAA unfoldases, we monitored how pArg^{AA} influences this interaction (Fig. 3c). A pull-down experiment using the NTD of ClpC (ClpC^{NTD}) revealed that pArg^{AA} inhibits the association between

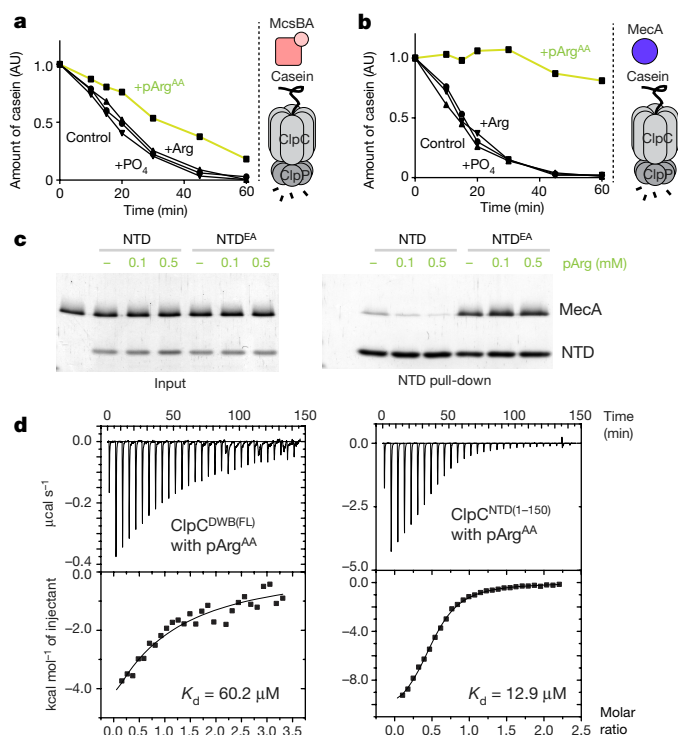


Figure 3 | Binding of pArg^{AA} to ClpC. **a, b**, Inhibitory effect of pArg^{AA} on McsBA-activated (**a**) and MecA-activated (**b**) ClpCP. AU, arbitrary units. **c**, Pull-down experiment monitoring the interaction of the NTD (wild type and E32A/E106A mutant) with MecA (10 μM each) in the presence pArg^{AA}. **d**, ITC profile of pArg^{AA} binding to full-length double Walker B mutant ClpC (ClpC^{DWB(FL)}; left) and the NTD of ClpC (ClpC^{NTD(1-150)}; right). Determined K_d values are indicated.

MecA and ClpC, probably by competing with MecA for the same binding site. When probing the direct interaction between pArg^{AA} and ClpC by isothermal titration calorimetry (ITC), we measured dissociation constants (K_d values) of 60 μM for full-length ClpC and 13 μM for the isolated NTD (Fig. 3d). The pronounced specificity for pArg^{AA} was confirmed by measuring the interaction with related compounds, which did not bind (pTyr^{AA}, arginine) or did so only weakly (phosphate). Moreover, we observed that pArg^{AA} did not bind to MecA, and has a very low affinity (K_d > 1 mM) for McsB (Extended Data Fig. 3). Finally, we tested the binding of pArg^{AA} to the NTD of ClpA, the closest homologue of ClpC in Gram-negative bacteria, which lack a protein arginine kinase. Because no binding was observed (Extended Data Fig. 3), the ability to recognise pArg^{AA} with high specificity seems to be a unique property of the ClpC unfoldase.

The pArg-binding site of ClpC could have two possible functions in protein degradation. It could serve as a docking site for the autophosphorylated form of McsB, which functions as an adaptor, or, alternatively, it could directly recognize pArg-containing substrates. To distinguish between these two possibilities, we enzymatically prepared pArg-modified β-casein (casein^{pArg}) (Fig. 4a and Extended Data Fig. 4). A pull-down assay showed that casein^{pArg}, but not unphosphorylated casein, binds to the ClpC^{NTD} (Fig. 4b). Because inhibition of the casein^{pArg}-NTD association required an excess of pArg^{AA}, it seems that a pArg residue in a protein context interacts more strongly with the NTD than the free phospho amino acid. When testing casein^{pArg} as a direct ClpCP substrate, we observed that the ClpCP protease could degrade casein^{pArg} even in the absence of McsB or MecA (Fig. 4c). These data suggest that the assembly of the functional ClpCP protease can proceed without adaptor proteins. To corroborate this surprising finding, we performed ClpC and ClpP pull-down experiments after incubation with substrate. Consistent with the degradation assays results, we observed the formation of a transient ClpCP complex in

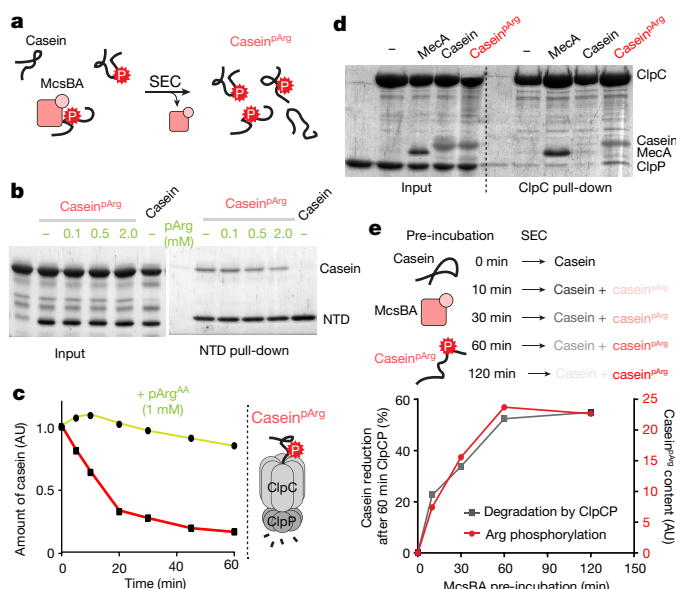


Figure 4 | ClpCP protease activity towards a pArg-containing substrate protein. **a**, Preparation of the casein^{pArg} model substrate. SEC, size exclusion chromatography. **b**, Binding of casein^{pArg} (35 μM) to ClpC^{NTD} (10 μM) at increasing amounts of pArg^{AA}. **c**, Degradation of casein^{pArg} by ClpCP without adaptor proteins and the inhibitory effect of pArg^{AA} on this activity. **d**, Pull-down experiment monitoring ClpCP complex formation in the presence of MecA, casein and casein^{pArg}. **e**, Preparation of substrate samples that contain increasing amounts of casein^{pArg} after prolonged incubation with McsBA. ClpCP degradation of resultant casein^{pArg} samples is directly correlated to the degree of substrate phosphorylation seen in pArg immunoblots (Extended Data Fig. 3e).

the presence of casein^{pArg}, even to a higher degree than in the presence of MecA. Conversely, unphosphorylated casein could not promote ClpCP assembly (Fig. 4d). Together, these data show that the pArg modification is crucial for recruiting substrates to the NTD of ClpC and for promoting assembly of the functional ClpCP protease complex.

To confirm the role of pArg as a degradation signal for the ClpCP protease, we analysed the digestion of substrate proteins that were arginine-phosphorylated to different degrees. For this purpose, we pre-incubated β-casein with McsB for increasing time intervals (Fig. 4e). ClpCP degradation assays of the resulting casein/casein^{pArg} mixtures showed a direct correlation between the amount of phosphorylated substrate and the extent of degradation (Fig. 4e and Extended Data Fig. 4e). Consistently, adding YwE phosphatase to the pArg-modified substrates abolished their degradation. These results unambiguously demonstrate that ClpCP selectively degrades casein^{pArg} and does not recognize pArg-less proteins as substrates.

The pArg docking sites of ClpC

To visualize how pArg binds to ClpC, we performed co-crystallization experiments of the NTD from *B. subtilis* ClpC with pArg^{AA}. The co-crystal structure was determined at 1.6 Å resolution (Extended Data Table 2) and, consistent with the symmetrical nature of the NTD protein fold²⁸ (Fig. 5a), revealed two almost identical pArg-binding sites (Fig. 5b, c). Mapping the electrostatic potential of the NTD on its molecular surface illustrates the ‘bipolar’ architecture of the pArg-binding site that distinguishes it from pSer/Thr or pTyr binding sites (Extended Data Fig. 5). Such organization is perfectly suited for simultaneously recognizing the positively charged guanidinium and the negatively charged phosphoryl group. The functional importance of the pArg recruitment is reflected by the exact conservation of the interacting residues in ClpC proteins from other Gram-positive species (Extended Data Fig. 6). Since the structural data suggest that the ClpC hexamer has 12 pArg-docking sites, we asked how many pArg tags per substrate are required for degradation. Native MS analysis of the

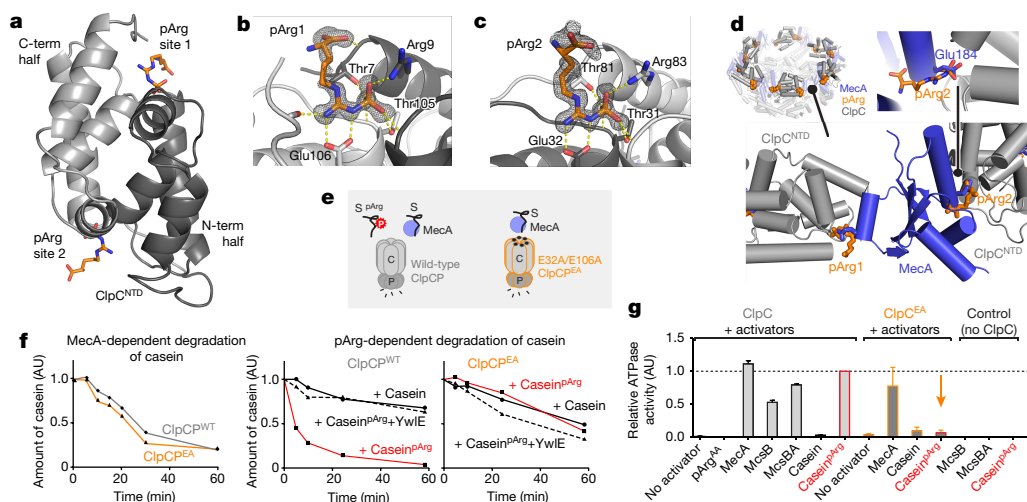


Figure 5 | Crystal structure of ClpC^{NTD} in complex with pArg. **a**, Overall structure of the ClpC^{NTD} domain bound to two pArg^{AA} molecules. The cartoon representation is coloured in light grey (residues 70–148) and dark grey (4–69) to highlight the two symmetrical halves of the NTD. **b**, **c**, Zoomed view of pArg-binding sites 1 (pArg1; **b**) and 2 (pArg2; **c**), with labelled interacting residues. The 2F_o – F_c omit electron densities of the pArg^{AA} ligands, calculated at 1.6 Å resolution, are contoured at 1σ. **d**, Overlap of MecA- and pArg-binding sites. Shown is the hexameric organization of the ClpC^{1–485} (grey) complex with MecA^{121–218} (blue) (PDB code 3PXG; ref. 29) superimposed with pArg^{AA} (orange). Bottom left, zoomed-in view shows two adjacent ClpC^{NTD} domains with pArg^{AA}

and MecA interactors. Bottom right, zoomed-in view illustrates that the pArg phosphoryl group and Glu184 (Glu198 in second binding site) of MecA compete for the same ClpC binding pocket. **e**, Scheme representing the distinct substrate (S) preferences of the wild-type ClpCP and ClpC^{EA} protease complexes. **f**, Degradation assays comparing the activity of wild-type ClpCP and ClpC^{EA} towards MecA-delivered (left) and pArg-labelled (middle, right) casein. YwIE was used as a control for the pArg-dependent degradation. **g**, ATPase activity of ClpC and ClpC^{EA} in the presence of putative substrate proteins. Levels are normalized to the induced ATPase activity of the ClpC–casein^{pArg} complex. Error bars show the s.d. of three independent experiments.

casein^{pArg} sample that resulted from prolonged incubation with McsB and that was completely degraded after ClpCP incubation revealed a mixture of mono- and di-phosphorylated molecules (Extended Data Fig. 7). Given the efficiency of also cleaving less phosphorylated substrates (Fig. 4f), we suppose that proteins carrying a single pArg mark can be degraded by ClpCP.

Notably, the identified pArg-binding sites match the MecA-binding grooves observed in the MecA–ClpC complex²⁹, with the phosphoryl moieties of pArg^{AA} binding in place of the MecA glutamate residues 184 and 198 (Fig. 5d). This overlap explains the inhibition of MecA–ClpCP by pArg^{AA}, as observed in our functional studies. Importantly, pArg^{AA} binds to ClpC glutamate residues 32 and 106, which in the MecA–ClpC complex remain unbound. Therefore, the bipolar architecture of the pArg-binding sites can only be fully explored by the phosphoguanidinium moiety, whereas the MecA glutamate residues only dock into the positively charged half. To test the structurally characterized binding mode experimentally, we prepared NTD mutants carrying an Glu32Ala/Glu106Ala double mutation (EA) and measured the interaction with MecA and pArg^{AA}. As predicted, mutating the two glutamate residues abolished pArg^{AA} binding (Extended Data Fig. 3) but did not impair the interaction with MecA (Fig. 3c). Consistent with the binding data, the corresponding ClpC^{EA} protease efficiently degraded protein substrates with the help of MecA, but failed to degrade substrates in a pArg-dependent manner (Fig. 5e, f). To confirm the selective failure in recognizing and degrading pArg protein, we measured the ClpC ATPase activity, which, in analogy to other AAA unfoldases³⁰, should be stimulated by substrates. We observed that purified casein^{pArg} could stimulate the ATPase activity of wild-type ClpC to a similar extent to MecA (Fig. 5g). In strong contrast, the ATPase of the ClpC^{EA} mutant could be induced by MecA, but not by casein^{pArg}, confirming the selective loss of pArg-dependent functions in this mutant.

Biological role of the pArg–ClpCP system

The developed ClpC^{EA} mutant represents a valuable tool for addressing the biological role of pArg-dependent protein degradation. To this end, we analysed the ability of the ClpC^{EA} mutant to suppress growth defects of a $\Delta clpC$ strain at increased temperatures (Fig. 6a). Whereas

the expression of wild-type ClpC from a plasmid restored thermotolerance and even made the bacteria more robust in surviving increased temperatures, expression of the ClpC^{EA} mutant could not rescue the bacteria. As the ClpC^{EA} mutant is fully functional as a protease and can team up with adaptor proteins, these data highlight the essential role of the pArg-dependent degradation pathway in surviving proteotoxic stress situations.

Discussion

Energy-dependent proteases are essential for all living organisms to carry out protein quality control and degrade short-lived regulatory proteins. In contrast to eukaryotes, which universally use polyubiquitin chains for marking target proteins, a general post-translational modification regulating proteolysis in bacteria is not known. Here, we characterize such a tagging system. We show that pArg is a degradation mark for the ClpCP proteolytic machine, present in most Gram-positive species. Despite differing in size, the bacterial pArg modifier shares several features with the eukaryotic polyubiquitin degradation tag. First, both pArg and polyubiquitin are post-translationally attached to substrates, allowing for dynamic regulation of degradation that is not available to mechanisms relying on sequence-encoded degrons. Second, the pArg mark is recognized by highly specific receptor sites on the NTD of ClpC (Fig. 6b), as is ubiquitin by special receptor proteins of the 19S regulatory particle. Third, owing to charge inversion, the phosphorylation of arginine residues is predicted to destabilize the native structure of substrate proteins, priming them for subsequent catalysed unfolding. Similarly, the polyubiquitin tag affects the structure and stability of marked proteins³¹. Fourth, the pArg tag is reversibly attached to substrate proteins. As enzymes building polyubiquitin chains are opposed by de-ubiquitinases, the activity of the McsB kinase is counteracted by the pArg-specific phosphatase YwIE, thus allowing for regulation of the pArg degradation pathway.

In addition to revealing the pArg degradation tag, our study clarifies the mechanism of the ClpCP protease. Notably, ClpC has been reported to be a unique AAA enzyme that requires accessory proteins to assemble its functional hexameric form³². The present data suggest that this model is not fully correct, as the degradation of pArg-containing

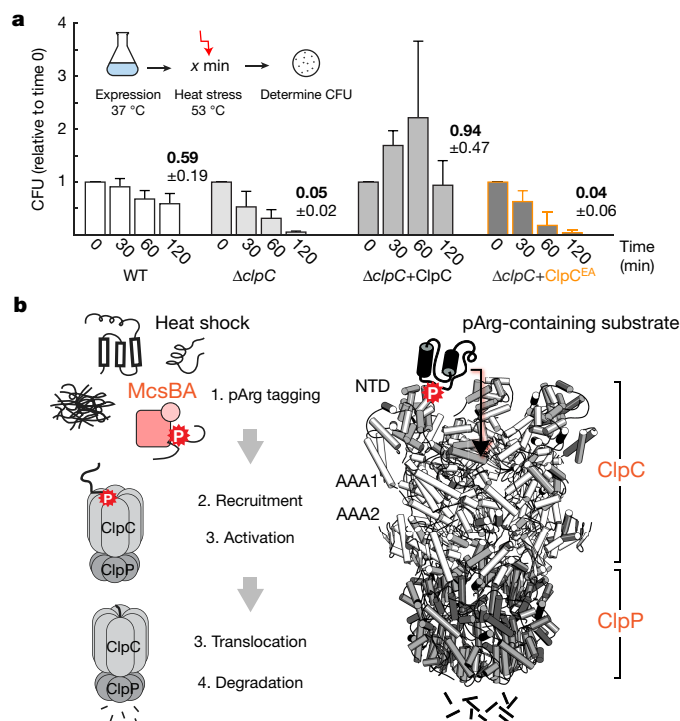


Figure 6 | The pArg-ClpCP degradation system. a, Thermotolerance assay analysing the *in vivo* complementation of $\Delta clpC$ by expressing ClpC and ClpC^{EA}. Levels are normalized to the values before heat shock (time 0), and numbers above bars represent the fraction of cells surviving after 2 h heat shock. CFU, colony-forming units. Error bars show the s.d. of three independent experiments. **b**, The pArg-ClpCP system. Left, cartoon representation shows that after phosphorylation by the McsB arginine kinase, pArg-tagged proteins are targeted to the ClpCP protease. Binding of pArg proteins to one of the 12 NTD binding pockets stimulates the ATPase activity of ClpC, leading to the translocation of the captured substrate into the ClpP protease cage and to protein degradation. Right, a model of the respective ClpCP complex (ClpC^{NTD} in light grey, ClpC AAA1/2 in white, ClpP in dark grey, substrate in black).

substrates does not require any additional co-factors. Substrate recruitment itself induces the ATPase activity of ClpC and promotes assembly of the functional ClpCP complex. Similar to the eukaryotic 26S proteasome, binding of specifically marked substrates is thus directly linked to protease activation. On the basis of the described functional similarities, the discovered pArg-ClpCP system seems to represent a simple bacterial version of the eukaryotic ubiquitin-proteasome system.

Our study also provides important insights into the biological role of pArg as a degradation tag. Analysis of the *in vivo* ClpP degradome suggests that the pArg tag is crucial not only for the regulatory proteolysis of CtsR but also for general turnover of structurally and functionally diverse proteins. Furthermore, we observed that pArg-dependent protein degradation is vital for coping with proteotoxic stress (Fig. 6a), and that pArg proteins are markedly enriched in the aggregate fraction of *B. subtilis* cells (Supplementary Tables 4 ($\Delta clpP$) and 5 (wild-type)). Of note, 50% of the *in vivo* phosphorylated proteins^{15,19} carry the phosphomark in a region predicted to adopt a defined secondary structure, that is, in an α -helix or β -strand. Presumably, these sites would only be accessible to McsB when present in an at least partially unfolded state, indicating that McsB might target the damaged form of those proteins. Consistent with this, a recent *in vivo* study points to the importance of the McsB kinase for removing aberrant proteins in *B. subtilis*: the deletion of the kinase led to the accumulation and aggregation of an unstable model protein, while the levels of a stably folded counterpart of this model protein were not influenced³³. We thus presume that protein arginine phosphorylation may have a role in the quality control of bacterial proteins, targeting unstable and aggregation-prone

proteins for ClpCP degradation. Modifying arginine of all amino acids to decide about the fate of aberrant proteins seems to make sense. As arginine-rich patches correlate with aggregation propensity³⁴, adding a phosphoryl group to arginine residues could hinder aggregation and, at the same time, promote the clearance of such problematic protein species by co-working protease machines.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2015; accepted 29 September 2016.

Published online 6 October 2016.

- Goldberg, A. L. The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. *Eur. J. Biochem.* **203**, 9–23 (1992).
- Löwe, J. et al. Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* **268**, 533–539 (1995).
- Glickman, M. H. et al. A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**, 615–623 (1998).
- Elsasser, S. & Finley, D. Delivery of ubiquitinated substrates to protein-unfolding machines. *Nat. Cell Biol.* **7**, 742–749 (2005).
- Kerscher, O., Felberbaum, R. & Hochstrasser, M. Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* **22**, 159–180 (2006).
- Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
- Burns, K. E., Liu, W. T., Boshoff, H. I., Dorrestein, P. C. & Barry, C. E., III. Proteasomal protein degradation in Mycobacteria is dependent upon a prokaryotic ubiquitin-like protein. *J. Biol. Chem.* **284**, 3069–3075 (2009).
- Pearce, M. J., Mintseris, J., Ferreyra, J., Gygi, S. P. & Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of *Mycobacterium tuberculosis*. *Science* **322**, 1104–1107 (2008).
- Battesti, A. & Gottesman, S. Roles of adaptor proteins in regulation of bacterial proteolysis. *Curr. Opin. Microbiol.* **16**, 140–147 (2013).
- Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T. & Baker, T. A. Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* **11**, 671–683 (2003).
- Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612 (2011).
- Keiler, K. C., Waller, P. R. & Sauer, R. T. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271**, 990–993 (1996).
- Fuhrmann, J. et al. McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR. *Science* **324**, 1323–1327 (2009).
- Krüger, E., Zühlke, D., Witt, E., Ludwig, H. & Hecker, M. Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *EMBO J.* **20**, 852–863 (2001).
- Schmidt, A. et al. Quantitative phosphoproteomics reveals the role of protein arginine phosphorylation in the bacterial stress response. *Mol. Cell. Proteomics* **13**, 537–550 (2014).
- Derré, I., Rapoport, G. & Msadek, T. CtsR, a novel regulator of stress and heat shock response, controls clp and molecular chaperone gene expression in gram-positive bacteria. *Mol. Microbiol.* **31**, 117–131 (1999).
- Kirstein, J., Dougan, D. A., Gerth, U., Hecker, M. & Turgay, K. The tyrosine kinase McsB is a regulated adaptor protein for ClpCP. *EMBO J.* **26**, 2061–2070 (2007).
- Elsholz, A. K., Michalik, S., Zühlke, D., Hecker, M. & Gerth, U. CtsR, the Gram-positive master regulator of protein quality control, feels the heat. *EMBO J.* **29**, 3621–3629 (2010).
- Elsholz, A. K. et al. Global impact of protein arginine phosphorylation on the physiology of *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **109**, 7451–7456 (2012).
- Boisvert, F. M., Chénard, C. A. & Richard, S. Protein interfaces in signaling regulated by arginine methylation. *Sci. STKE* **2005**, re2 (2005).
- Feng, J. et al. Trapping and proteomic identification of cellular substrates of the ClpP protease in *Staphylococcus aureus*. *J. Proteome Res.* **12**, 547–558 (2013).
- Miethe, M., Hecker, M. & Gerth, U. Involvement of *Bacillus subtilis* ClpE in CtsR degradation and protein quality control. *J. Bacteriol.* **188**, 4610–4619 (2006).
- Derré, I., Rapoport, G. & Msadek, T. The CtsR regulator of stress response is active as a dimer and specifically degraded *in vivo* at 37 degrees C. *Mol. Microbiol.* **38**, 335–347 (2000).
- Fuhrmann, J. et al. Structural basis for recognizing phosphoarginine and evolving residue-specific protein phosphatases in Gram-positive bacteria. *Cell Reports* **3**, 1832–1839 (2013).
- Trentini, D. B., Fuhrmann, J., Mechtler, K. & Clausen, T. Chasing phosphoarginine proteins: development of a selective enrichment method using a phosphatase trap. *Mol. Cell. Proteomics* **13**, 1953–1964 (2014).
- Schlothauer, T., Mogk, A., Dougan, D. A., Bukau, B. & Turgay, K. MecA, an adaptor protein necessary for ClpC chaperone activity. *Proc. Natl Acad. Sci. USA* **100**, 2306–2311 (2003).

27. Kirstein, J., Zühlke, D., Gerth, U., Turgay, K. & Hecker, M. A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in *B. subtilis*. *EMBO J.* **24**, 3435–3445 (2005).
28. Kojetin, D. J. *et al.* Structural and motional contributions of the *Bacillus subtilis* ClpC N-domain to adaptor protein interactions. *J. Mol. Biol.* **387**, 639–652 (2009).
29. Wang, F. *et al.* Structure and mechanism of the hexameric MecA-ClpC molecular machine. *Nature* **471**, 331–335 (2011).
30. Woo, K. M., Kim, K. I., Goldberg, A. L., Ha, D. B. & Chung, C. H. The heat-shock protein ClpB in *Escherichia coli* is a protein-activated ATPase. *J. Biol. Chem.* **267**, 20429–20434 (1992).
31. Hagai, T. & Levy, Y. Ubiquitin not only serves as a tag but also assists degradation by inducing protein unfolding. *Proc. Natl Acad. Sci. USA* **107**, 2001–2006 (2010).
32. Kirstein, J. *et al.* Adaptor protein controlled oligomerization activates the AAA+ protein ClpC. *EMBO J.* **25**, 1481–1491 (2006).
33. Stannek, L., Gunka, K., Care, R. A., Gerth, U. & Commichau, F. M. Factors that mediate and prevent degradation of the inactive and unstable GudB protein in *Bacillus subtilis*. *Front. Microbiol.* **5**, 758 (2015).
34. Chan, P., Curtis, R. A. & Warwicker, J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Sci. Rep.* **3**, 3333 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank E. Charpentier for providing *B. subtilis* strains and plasmids, S. Spiess for advice on *B. subtilis* genetic manipulation,

G. Dürnberger for assistance with statistical analysis, A. Schmidt for help with mass spectrometry interpretation, R. Beveridge for performing native mass spectrometry analysis, J. Kley for help with ITC experiments and G. Bourenkov and T. Schneider at DESY for assistance during data collection. This work was supported by the Austrian Research Promotion Agency (FFG). The IMP is funded by Boehringer Ingelheim.

Author Contributions D.B.T. performed mass spectrometry and biochemical experiments; M.J.S. performed the structural analysis and biochemical experiments; A.H. and R.K. performed *in vivo* complementation assays and biochemical experiments; L.D. assisted with biochemical experiments; T.C. designed the study with expert assistance from K.M. for the mass spectrometry part; D.B.T., M.J.S. and T.C. wrote the manuscript, with contributions from all authors.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession code 5HBN. The mass spectrometry data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) with the dataset identifier PXD003305. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.C. (clausen@imp.univie.ac.at).

Reviewer Information *Nature* thanks S. Gygi, Y. Shi and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

DNA construct design. For recombinant protein overexpression in *Escherichia coli*, the *clpP*, *mcsB*, *mcsC*, *clpC* and *clpC*¹⁻¹⁵⁰ (NTD) genes/fragments from *B. subtilis* and *clpA*¹⁻¹⁵⁰ (NTD) from *E. coli* were cloned into pET21 or pET SUMO vectors (Novagen) conferring a terminal hexahistidine (6His) tag. Previously published pET21-derived plasmids were used for the expression of *Geobacillus stearothermophilus* McsB, YwIE and YwIE^{D118N} proteins^{13,24,25}. For protein expression in *B. subtilis*, genes were cloned into the vector pHCMC05 (ref. 35) that contains the Psac IPTG-inducible promoter. For the expression of the ClpP, ClpP^{TRAP} (S98A), ClpP^X (E119A/R142E), ClpP^{X-TRAP} (S98A/E119A/R142E), ClpC and ClpC^{EA} (E32A/E106A), a C-terminal 6His tag including a Leu-Glu linker was introduced by PCR amplification. Single point mutations were generated using the QuikChange II mutagenesis kit (Agilent Technologies).

***B. subtilis* strains.** The pHCMC05 plasmids described above were transformed into wild-type *B. subtilis* (strain 168) (ATCC 2385). To generate ClpC-knockout ($\Delta clpC$) bacteria, genomic DNA from a *B. subtilis* $\Delta clpC::tet$ strain³⁶ was transformed into the *B. subtilis* strains containing the pHCMC05, pHCMC05-clpP^X, pHCMC05-clpP^{X-TRAP}, pHCMC05-ClpC or pHCMC05-ClpC-E32A-E106A plasmids. The disruption of *clpC* in the resulting strains was confirmed by sequencing. A previously published *B. subtilis* $\Delta clpP::spec^R$ (ref. 37) strain was used to grow ClpP-knockout bacteria for phosphoproteomic analysis.

Culture procedures. All bacterial cultures were grown in Luria-Bertani (LB) medium. For the *B. subtilis* $\Delta clpC$ and $\Delta clpP$ strains, tetracycline (10 $\mu\text{g ml}^{-1}$) and spectinomycin (100 $\mu\text{g ml}^{-1}$) were added, respectively. *E. coli* and *B. subtilis* cultures containing the pHCMC05-derived plasmids were cultured in the presence of ampicillin (50 $\mu\text{g ml}^{-1}$) and chloramphenicol (10 $\mu\text{g ml}^{-1}$), respectively. *E. coli* BL21 (DE3) containing pET21- or pET SUMO-derived vectors were cultured in the presence of ampicillin (50 $\mu\text{g ml}^{-1}$).

In vivo pull-downs using ClpP trapping mutants. For the ClpP^{X-TRAP} pull-down experiment in the wild-type background of *B. subtilis* (Fig. 1b), 6 independent *B. subtilis* cultures were grown in LB media expressing either ClpP^X (3 control cultures to identify unspecific binding partners) or ClpP^{X-TRAP} (3 sample cultures). After the cells were grown at 37 °C until mid-exponential phase, expression of His-tagged ClpP^X or ClpP^{X-TRAP} proteins was induced with 1 mM IPTG. Recombinant protein expression proceeded for 3 h at 37 °C. To induce the activity of heat-shock proteins, including McsB and various Clp ATPases, the cultures were incubated in a pre-warmed incubator at 45 °C for 45 min. Cells were collected by centrifugation, resuspended in lysis buffer (25 mM Tris, pH 7.5, 150 mM NaCl, 10% glycerol) and stored at -80 °C. For the pull-down experiment comparing ClpP-trapped proteins in wild-type and $\Delta clpC$ backgrounds (Fig. 1c), the same procedure was applied using 12 independent *B. subtilis* cultures (3 wild type with ClpP^X, 3 wild type with ClpP^{X-TRAP}, 3 $\Delta clpC$ with ClpP^X, 3 $\Delta clpC$ with ClpP^{X-TRAP}).

The thawed cell suspensions were incubated for 1 h on ice with 2 mg ml⁻¹ lysozyme (Sigma), Complete protease inhibitor cocktail (Roche), 0.2 mM PMSF (Sigma) and 10 $\mu\text{g ml}^{-1}$ DNase (Sigma). Cells were sonicated and the resultant lysate was cleared by centrifugation at 4 °C. For the purification of 6His-tagged ClpP, the lysate was incubated with Dynabeads His-Tag Isolation & Pulldown (Invitrogen) for 1 h at 4 °C. The beads were then washed 5 \times with lysis buffer and 2 \times with lysis buffer containing 50 mM imidazole. Two aliquots (5 or 10%) of the resulting beads were collected for SDS-PAGE analysis (protein elution with denaturing SDS-PAGE sample buffer) or Tris-acetate native-PAGE (protein elution with lysis buffer containing 500 mM imidazole). The remaining beads were subjected to reduction with 2 mM DTT (56 °C, 40 min), alkylation with 10 mM iodoacetamide (room temperature, in the dark, 45 min), and digestion with 2.5 μg Trypsin Gold (Promega) at 37 °C for 12 h.

MALDI-MS analysis of ClpP purifications. To analyse the mass of *B. subtilis* proteins co-purifying with ClpP (6His) under heat-shock conditions (Extended Data Fig. 1), matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was performed. The corresponding protein purifications were spotted on a MALDI plate using a sinapinic acid (10 mg ml⁻¹) matrix prepared in 50% acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA). The samples were measured in a 4800 MALDI-TOF-TOF (AB Sciex) instrument operated in linear mode. Calibration was performed internally using cytochrome c as standard.

Sample preparation for phosphoproteomic analysis. For the phosphoproteomic analysis of total cell extracts of *B. subtilis* $\Delta clpC$, cells were lysed by sonication in buffer 4% SDS, 100 mM Tris, pH 7.5, 100 mM DTT and further processed using a filter aided sample preparation (FASP)³⁸ modified method, as described previously¹⁵, followed by trypsin digestion at 37 °C for 12–16 h. Protein aggregates were dissolved in the SDS buffer and processed in the same manner. Trypsin digestion completion was inspected by retention time and UV intensity (214 nm) distribution upon reverse-phase high-performance liquid chromatography (RP-HPLC) separation of a 0.1% aliquot of the resulting supernatants on a monolithic column

(Ultimate Plus equipped with a PepSwift PS-DVB column, 5 cm \times 200 μm , Dionex-Thermo-Fisher).

For the ClpP *in vivo* pull-down assays, a small aliquot (0.5%) of the on-bead trypsin digests was collected for subsequent quantitative analyses of co-purified proteins. The biological replicates were then pooled and further processed for phosphorylation analysis. Before phosphopeptide enrichment, sample digests were purified from buffer reagents by RP-C18 solid phase extraction at neutral pH using Oasis HLB cartridges (Waters). A previously described TiO₂ protocol¹⁵, optimized in accordance to the acid-labile nature of phosphoarginine, was used for phosphopeptide enrichment.

LC-MS/MS analysis. Reverse-phase separation of all peptide mixtures was carried out on an Ultimate 3000 RSLC nano-flow chromatography system (Thermo Scientific), using 0.5% acetic acid (pH 4.5 with NH₃) as loading solvent, to prevent phosphoarginine hydrolysis during removal of salts in the pre-column (PepMapAcclaim C18, 5 mm \times 0.3 mm, 5 μm , Thermo Scientific). Peptide separation was achieved on a C18 separation column (PepMapAcclaim C18, 50 cm \times 0.75 mm, 2 μm , Thermo Scientific) by applying a linear gradient from 2% to 35% solvent B (80% ACN, 0.08% formic acid) in 120 or 240 min (pull-down and total extract samples, respectively) at a flow rate of 230 nl min⁻¹. Solvent A was 2% ACN, 0.1% formic acid. The separation was monitored by UV detection and the outlet of the detector was directly coupled to the nano-electrospray ionization source (Proxeon Biosystems) for MS analysis.

For phosphorylation analysis, TiO₂ elution samples were infused into the LTQ Orbitrap Velos Pro ETD mass spectrometer (Thermo Scientific) using PicoTip nanospray emitter tips (New Objective) at a voltage of 1.5 kV. Peptides were analysed in data-dependent fashion in positive ionization mode, applying two different fragmentation methods: collision-induced dissociation (CID) and electron-transfer dissociation (ETD). The survey scan was acquired at resolution 60,000 and the 6 most abundant signals with charge state equal or higher than 2+ and exceeding an intensity threshold of 1,500 counts were selected for peptide fragmentation analysis. For MS/MS experiments, precursor ions were isolated within a 2.1-Da window centred on the observed *m/z*. To prevent repeated fragmentation of highly abundant peptides, selected precursors were dynamically excluded for 30 s from MS/MS analysis. CID fragmentation was achieved at normalized collision energy (NCE) of 35% with additional activation of the neutral loss precursor at M-49, M-32.7 and M-98 AMU in a standard multistage activation method. For ETD, peptides were incubated with fluoranthene anions allowing for charge-state-dependent incubation times (90 ms for 3+ charged peptides), and resulting peptide fragments were detected in the ion trap analyser.

For the identification of co-purified proteins in the ClpP *in vivo* pull-down assays, slightly different instrument settings were used. The 12 most abundant signals with charge state equal to or higher than 2+ and exceeding an intensity threshold of 500 counts were selected for CID peptide fragmentation analysis, applying an isolation window of 2 Da. Multistage activation was disabled. Selected precursors were dynamically excluded for 60 s from MS/MS analysis. Each pull-down digest was analysed twice, to evaluate technical reproducibility.

MS data analysis. For the phosphorylation analysis of TiO₂-enrichment samples, raw data were extracted by the Proteome Discoverer software suite (version 1.4.0.288, Thermo Scientific) and searched against a combined forward/reversed database of *B. subtilis* Uniprot Reference Proteome with common contaminants added (4,455 entries in total) using MASCOT (version 2.2.07, Matrix Science). Carbamidomethylation of cysteine was set as fixed modification. Phosphorylation of serine, threonine, tyrosine and arginine plus oxidation of methionine were selected as variable modifications. Since tryptic cleavage is impaired at phosphorylated arginine, a maximum of two missed cleavage sites was allowed, whereas fully tryptic cleavage of both termini was required. The peptide mass deviation was set to 5 p.p.m.; fragment ions were allowed to have a mass deviation of 0.8 Da. False discovery rates were assessed using the Percolator tool³⁹ within the Proteome Discoverer package. The results were filtered for peptide rank 1 and high identification confidence, corresponding to a 1% false discovery rate. Low-scoring peptides (Mascot ion score ≤ 20) were manually verified. In the rare cases in which a peptide was mapped to more than one protein sequence, both protein hits are reported. For reliable phosphorylation site analysis, all phosphopeptide hits were automatically re-analysed by the phosphoRS software⁴⁰ within the Proteome Discoverer software suite. All the phosphopeptides identified in the ClpP *in vivo* pull-down assays were manually inspected. For other samples, we considered a phosphorylation site to be localized when the reported phosphoRS probability was higher than 90%. When multiple peptide-spectrum matches (PSMs) were obtained for the same phosphopeptide, only the PSM presenting the best identification/localization score compromise is presented. The multiple redundant PSMs were ranked according to their phosphoRS probability score into three categories (90–94%, 94–97% and 97–100%); the PSM presenting the best Mascot score within the highest phosphoRS category achieved was reported. PSMs presenting wrong or inconclusive

localizations were thus excluded from the final list of phosphopeptides. Multiply phosphorylated peptides were also excluded from the analysis, because they cannot be classified into 'phosphorylation type' categories.

For quantitative analysis of ClpP-trapped proteins, MS data were analysed using the MaxQuant software environment⁴¹, version 1.5.1.2, and its built-in Andromeda search engine⁴², against the *B. subtilis* Uniprot database described above. Strict trypsin specificity with up to two missed cleavages was used. The minimum required peptide length was set to six amino acids. Carbamidomethylation of cysteine was set as a fixed modification and *N*-acetylation of proteins N termini (42.010565 Da) and oxidation of methionine were set as variable modifications. During the main search, parent masses were allowed an initial mass deviation of 4.5 p.p.m. and fragment ions were allowed a mass deviation of 0.5 Da. The mass accuracy of the precursor ions was improved by time-dependent recalibration algorithms of MaxQuant. The 'match between runs' option was enabled to match identifications across samples within a time window of 2 min of the aligned retention times. The second peptide identification option in Andromeda was enabled. PSM and protein identifications were filtered using a target-decoy approach at false discovery rate of 1% for PSMs and 5% for proteins. Relative, label-free quantification of proteins was done using the MaxLFQ algorithm⁴³ integrated into MaxQuant using default parameters. Unique and razor peptides were considered for quantification.

Statistical evaluation of the resulting protein quantifications was performed using R scripting. Proteins quantified in less than 50% of the samples were filtered out. Missing LFQ values were substituted by the lowest value observed in the corresponding sample. For each protein, the fold change of LFQ-averaged intensities (ratio ClpP^{TRAP}/control) and the corresponding *P* value (Limma test; Linear Models for Microarray Data) were calculated. A protein was considered to be a ClpP substrate when it was found to be enriched in the TRAP pull-down assays by a factor of at least 2 and *P* < 0.05. The 'protein groups' output file from MaxQuant containing the statistical evaluation is available in Supplementary Table 1. The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository⁴⁴ (<http://proteomecentral.proteomexchange.org>). Representative spectra of the pArg peptides identified in the ClpP trapping mutant pull-downs are presented in Supplementary Fig. 1.

Expression and purification of recombinant proteins. For the overexpression of recombinant proteins in *E. coli* BL21 (DE3), LB cultures were grown at 37 °C until the exponential phase, when expression was induced with 0.5 mM IPTG. After expression, cells were collected by centrifugation, resuspended in buffer A and stored at -80 °C. As the *B. subtilis* ClpC protein was unstable when expressed in *E. coli*, the production of wild-type ClpC(6His) and ClpC^{EA}(6His) was performed in *B. subtilis* containing the corresponding pHCMC05 plasmids, by induction with 1 mM IPTG for 3 h at 37 °C. The optimal expression strategies and purification buffers are summarized in Extended Data Table 3.

Cell suspensions were incubated on ice for 30 min in the presence of 1 mg ml⁻¹ lysozyme, 0.1 mM PMSF, 10 µg ml⁻¹ DNase and sonicated. Lysates were cleared by centrifugation and loaded on a 5 ml Ni- or Co-NTA column (GE Healthcare LifeSciences) equilibrated in buffer A. Washes were performed using a step-wise imidazole gradient, typically starting with 25 mM. The His-tagged proteins were eluted with buffer A containing 250 mM imidazole and concentrated using Vivaspins (Sartorius Stedim Biotech). Constructs expressed as a SUMO-fusion (SUMO-ClpP and SUMO-MecA) were incubated with SUMO Protease (Thermo Fisher Scientific) to obtain tag-free versions of the proteins. All resulting proteins were further purified by gel filtration on a Superdex-75 or -200 column (GE Healthcare LifeSciences) equilibrated with buffer B. For the purification of McsA(6His), the full-length protein was separated from an abundant cleavage product by ion exchange on a MonoQ column (GE Healthcare LifeSciences) using a 0.1–1 M NaCl gradient in 50 mM Tris, pH 8.5, 1 mM TCEP. All proteins were aliquoted and stored at -80 °C until further use.

For the purification of the *B. subtilis* ClpC^{NTD} and NTD^{EA} mutant, affinity, ion exchange, and size exclusion chromatography were carried out in a 0.5× PBS buffer (6 mM Na/K phosphate, pH 7.25, 1.35 mM KCl, 68.5 mM NaCl). After elution from Ni-NTA using the PBS buffer supplemented with 250 mM imidazole, the protein was passed through a ResourceQ column (GE Healthcare LifeSciences). This was followed by gel filtration on a Superdex 200 column (GE Healthcare LifeSciences). **In vitro pull-down assays.** Varying amounts of the analysed components (individual proteins, pArg^{AA}) were incubated in 200 µl of reaction buffer (10 mM Tris, pH 8.0, 50 mM NaCl, 5 mM MgCl₂ and 25 mM imidazole). After 5 min, 50 µl of Ni-Sepharose was added to capture His-tagged proteins. Sepharose was washed three times with 200 µl reaction buffer and bound proteins were eluted with 50 µl of elution buffer (10 mM Tris, pH 8.0, 50 mM NaCl, 5 mM MgCl₂ and 1 M imidazole). For visualization, input and elution, fractions were analysed by SDS-PAGE. **Native-PAGE.** For the analysis of the oligomeric state of purified ClpP samples, a Tris-acetate non-denaturing PAGE system was used. Gel composition

corresponded to 7, 10 or 15% acrylamide, 0.24% bis-acrylamide, 200 mM Tris-acetate, pH 7, polymerized in the presence of 0.042% ammonium persulfate (APS) and 0.125% N,N,N',N'-tetramethylethylenediamine (TEMED). The running buffer composition was 25 mM Tris-HCl, 192 mM glycine, pH 8.3. Protein separation was performed at 4 °C and 150 mV for 3–4 h, and proteins were visualized by Coomassie-based InstantBlue protein stain (Expedeon). The NativeMark unstained protein standard (Thermo Fisher Scientific) was used for the estimation of ClpP oligomeric state.

Purification of casein^{pArg}. To produce the arginine-phosphorylated substrate, β-casein from bovine milk (Sigma) was incubated at 10 µM (all protein concentrations, if not otherwise mentioned, are for a single protomer) with 2 µM McsB and McsA from *B. subtilis* in 25 mM Tris, pH 7.5, 50 mM NaCl, 20 mM MgCl₂, and 5 mM ATP at 30 °C for 2 h. The reaction mixture was concentrated with a Vivaspin device and applied to a Superdex-200 size exclusion column equilibrated with 25 mM Tris, pH 7.5, 50 mM NaCl. The fractions most strongly enriched in β-casein over McsB (Extended Data Fig. 5a) were then pooled, concentrated and stored at -80 °C. The presence of the pArg modification was confirmed by immunoblotting using a pArg-specific antibody as described below.

To obtain casein^{pArg} of higher purity, a similar phosphorylation reaction was performed, except that 4 µM of the *G. stearothermophilus* McsB(6His) was used instead of the *B. subtilis* McsBA complex. The reaction mixture was afterwards applied to a Ni-NTA column (GE Healthcare LifeSciences) to reduce the amounts of the His-tagged McsB before the final gel filtration purification step (Extended Data Fig. 5c).

To prepare β-casein phosphorylated to different degrees (Fig. 4e), a large-scale phosphorylation reaction was set-up with the *B. subtilis* McsBA at 30 °C. An aliquot (time 0) was collected before the addition of ATP. After adding ATP, aliquots were taken after 10, 30, 60 and 120 min. After adding 100 mM EDTA (to stop phosphorylation), each aliquot was concentrated by Vivaspin ultrafiltration and applied to a Superdex 200 size exclusion column. An additional sample, which was collected after 120 min incubation with McsBA, was treated with 1 µM YwIE arginine phosphatase for 2 h. The phosphatase was then inactivated by adding 2 mM pervanadate and the sample was concentrated and submitted to size exclusion chromatography. The different casein^{pArg} preparations were concentrated and stored at -80 °C.

To quantify the increase of arginine phosphorylation over time, a western blot was performed using the pArg-specific antibody described previously²⁴. The casein^{pArg} preparations (2 µg) were separated by SDS-PAGE, transferred to a nitrocellulose membrane and fixed using a 0.4% formaldehyde solution in PBS, pH 7.5, for 30 min as previously described⁴⁵. After blocking, the pArg-specific primary antibody (2 µg ml⁻¹, Morphosys AG) was incubated overnight at 4 °C and the secondary antibody (goat anti human IgG F(ab')₂:HRP, AbD Serotec) was used at a 1:7,000 dilution for 1.5 h at room temperature. The detection was performed using ECL Plus western blotting substrate (Pierce). The signals were quantified using the ImageJ software⁴⁶ and normalized to the band intensities observed in a coomassie-stained SDS-PAGE gel replicate having 1.3 µg of each protein preparation.

ClpCP in vitro degradation assays. In vitro degradation assays containing 0.16 µM ClpC (hexamer), 0.16 µM ClpP (heptamer) and 5 µM β-casein substrate were performed in 25 mM Tris, pH 7.5, 150 mM NaCl, 20 mM MgCl₂ and 5 mM ATP at 30 °C. A different β-casein concentration (10 µM) was used, for better resolution, when comparing ClpCP and ClpCP^{EA} in MecA-dependent degradation. Small molecule compounds, for example, phospho-L-arginine (pArg^{AA}, Toronto Biochemicals), L-arginine (Sigma) or sodium phosphate, pH 7.5, were added at 1 mM. Time-point aliquots were mixed with denaturing SDS-PAGE sample buffer containing 100 mM EDTA to stop the reaction and analysed by SDS-PAGE. The resulting gels were stained with InstantBlue dye (Expedeon) and quantified using ImageJ⁴⁶. Supplementary Fig. 2 shows full SDS-PAGE gels and corresponding quantifications of Figs 2–5.

Radioactive kinase assay. The radioactive kinase assay was performed at 19 °C in 25 mM Tris, pH 7.5, 50 mM NaCl, 2% glycerol, and 20 mM MgCl₂. 15 µM β-casein substrate was incubated with 1 µM McsB and/or other proteins (1 µM McsA, 5 µM YwIE) for 0, 30 and 120 min. The reaction was started by adding ATP (spiked with [^γ-³²P]ATP from Perkin Elmer) to a final concentration of 10 mM and stopped with denaturing SDS-PAGE sample buffer. After resolving the samples by SDS-PAGE, phosphorylation was visualized using phosphorimager technology (GE Healthcare LifeSciences).

ATPase assays. ATPase activity was determined by a coupled enzymatic reaction⁴⁷. 0.125–0.5 µM ClpC was incubated with 18.75 U ml⁻¹ pyruvate kinase, 21.45 U ml⁻¹ lactate dehydrogenase, 0.2–0.3 mM NADH, 7.5 mM phosphoenolpyruvate and 2 mM ATP in 20 mM HEPES, pH 7.5, 100 mM NaCl, and 5 mM MgCl₂. Further assay proteins (MecA, McsB, McsA and β-casein) were added at 4–6-fold excess over ClpC. The absorption at 340 nm (A_{340 nm}) was recorded for 60 min using a Synergy H1 Multi-Mode Reader. The molar ATPase activity (v) was calculated

by the equation: $v = \Delta A_{340\text{ nm}} / (\text{path length} \times 6,220 \times [\text{ClpC}] \times \text{M}^{-1} \times \text{cm}^{-1})$. All activity data represent a minimum of three independent experiments and the variability is highlighted as standard deviation.

ITC. ITC measurements were performed using VP-ITC (Microcal). Ligands (pArg^{AA}, L-arginine, phospho-L-tyrosine (pTyr^{AA}, from Sigma), and sodium phosphate, pH 7.5, were prepared at 0.3–1.4 mM in 25 mM Tris, pH 8, 50 mM NaCl, 0.2 mM TCEP and titrated to a 20 μM ClpC^{DWB} (full-length) or 140 μM ClpC^{NTD} protein solution present in the same buffer. The same set-up was used for the analysis of ClpA^{NTD}. The following settings were used: 5 μl (first) and 10 μl (all subsequent injections) injection volume, 300 s spacing time between the injections, 300 r.p.m. stirring speed, 25 °C temperature and overflow mode. Control experiments were carried out to correct for dilution effects upon protein/ligand titration. Resulting data were analysed with the MicroCal ORIGIN software.

Structure determination of *B. subtilis* ClpC^{NTD} in complex with pArg^{AA}. Crystals of a ClpC^{NTD}–pArg^{AA} complex were obtained by the sitting-drop vapour diffusion method upon mixing 100 nl of reservoir with 200 nl of a ClpC^{NTD} (2 mM) protein solution containing 2 mM pArg^{AA}. The optimized crystallization solution contained 13.5% (w/w) polyethylene glycol 4000, 500 mM ammonium sulfate, and 100 mM sodium acetate at pH 5. Crystals formed overnight at 19 °C and were soaked/cryo-protected in 40% polyethylene glycol 400, 20 mM Tris pH 8, and 6 mM pArg^{AA} before being flash-frozen. Diffraction data to 1.6 Å were collected at 100 K using a wavelength of 0.9763 Å at beamline P14, DESY, Hamburg and integrated with XDS⁴⁸. Molecular replacement in Phaser⁴⁹ using ClpC^{NTD} from *B. subtilis* (PDB code 2Y1Q; ref. 29) as a search model yielded a high-confidence solution. Refinement in CNS⁵⁰, automatic rebuilding in Phenix⁵¹, and manual rebuilding in Coot^{52,53} were carried out, followed by placing of ligands in Coot⁵⁴. *N*-(omega)-phospho-L-arginine structure and constraints were obtained with the respective SMILES code using eLBOW⁵⁵. Rounds of refinement in Phenix⁵⁶ and rebuilding in Coot yielded the final model with good statistics and geometry (Extended Data Table 2 and the following Ramachandran statistics: 98% favoured, 2% allowed, 0% outliers, 0% rotamer outliers). The featured-enhanced map, which is based on a composite residual omit map, was used to show ligand density. Figures were produced in PyMol⁵⁷.

Purification of *B. subtilis* protein aggregates. For the phosphoproteomic analysis of *B. subtilis* protein aggregates, a 3-l culture of *B. subtilis* was grown at 37 °C until late exponential phase and then heat-shocked (50 °C) for 45 min. Cells were collected by centrifugation, resuspended in 25 mM Tris, pH 7.5, 150 mM NaCl, 0.5% Triton X-100 and stored at –80 °C.

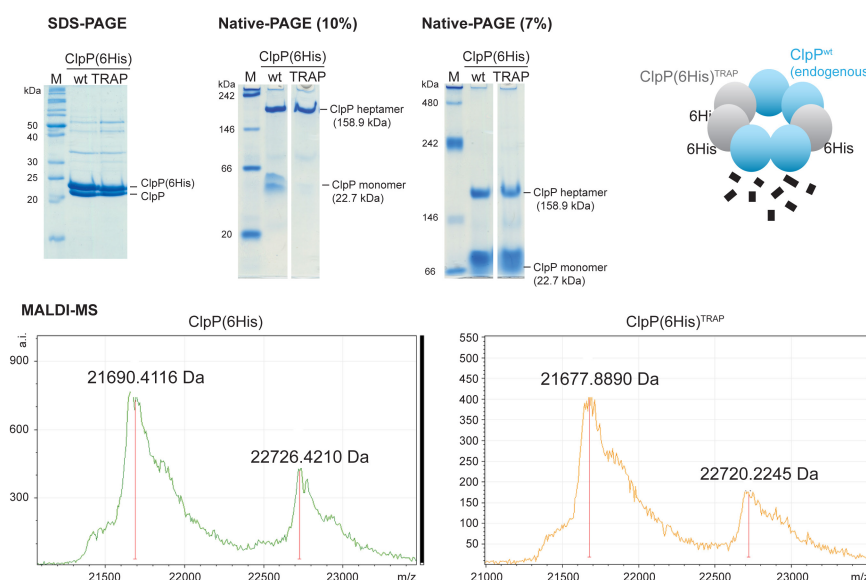
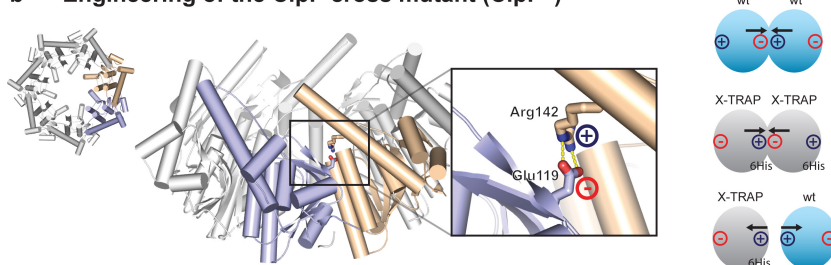
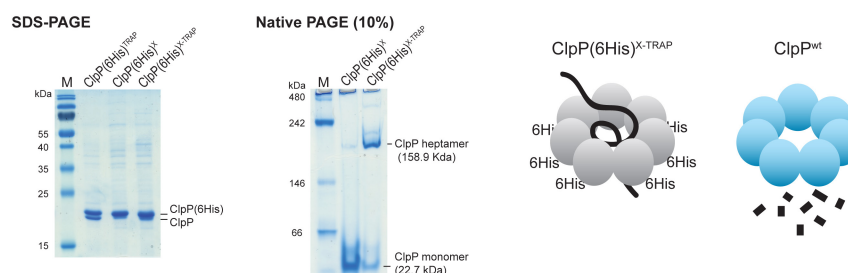
A 30-ml cell suspension in 25 mM Tris, pH 7.5, 150 mM NaCl, 0.5% Triton X-100 was incubated on ice with 3 mg ml^{–1} lysozyme, Complete protease inhibitor cocktail (Roche), 20 μg ml^{–1} DNase, 0.2 mM PMSF and 2 mM vanadate for 30 min. After dilution to 100 ml, cells were gently lysed at 4 °C by French Press (Constant Cell Disruption Systems) at 1.7 kbar. Lysis efficiency, estimated by plating out serial dilutions, exceeded 99%. The lysates were centrifuged at 45,000g for 30 min. The resulting pellets, containing insoluble protein aggregates, were further washed with 20 ml lysis buffer containing 0.4 mg ml^{–1} lysozyme, 10 μg ml^{–1} DNase, 0.2 mM PMSF and 2 mM vanadate. After 40 min homogenization at 4 °C under gentle agitation, the pellets were re-centrifuged. The protein aggregates contained in the pellets were then solubilized in 7 ml 25 mM Tris, pH 7.5, 8 M urea by sonication. The samples were again centrifuged to separate the urea-solubilized protein aggregates from cell debris. The resulting supernatants were stored at –80 °C until MS sample processing.

Thermotolerance assay. To test the role of ClpC during heat stress, the following *B. subtilis* strains were investigated: wild type + pHCMC05, $\Delta\text{clpC}::\text{tet}$ + pHCMC05, $\Delta\text{clpC}::\text{tet}$ + pHCMC05–ClpC and $\Delta\text{clpC}::\text{tet}$ + pHCMC05–ClpC^{EA}. For all experiments, cultures were grown at 37 °C in LB media containing 10 μg ml^{–1} chloramphenicol and 0.2 mM IPTG. After reaching exponential phase, the cultures were transferred to a pre-warmed incubator at 53 °C for 0, 30, 60 or 120 min, respectively. After heat stress, the samples were diluted sequentially and transferred to LB plates. To compare the survival rate after heat stress, we determined the number of colony-forming units (CFU). All experiments were independently performed three times and the observed variability is highlighted as standard deviation.

Native mass spectrometry. Native mass spectrometry experiments were carried out on a Synapt G2Si instrument (Waters) with a nano-electrospray ionization (nESI) source. Mass calibration was performed by a separate infusion of NaI cluster ions. Solutions were ionised through a positive potential applied to metal-coated borosilicate capillaries (Thermo Scientific). β -casein samples (5 μM) were sprayed from 25 mM ammonium acetate, pH 6.8. The temperature settings were capillary voltage 1.5 kV, sample cone voltage 30 V, extractor source offset 46 V, and source temperature 50 °C. Data were processed using Masslynx V4.1 software.

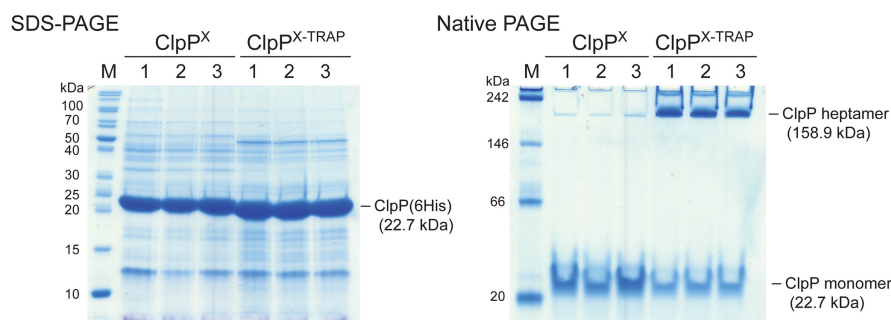
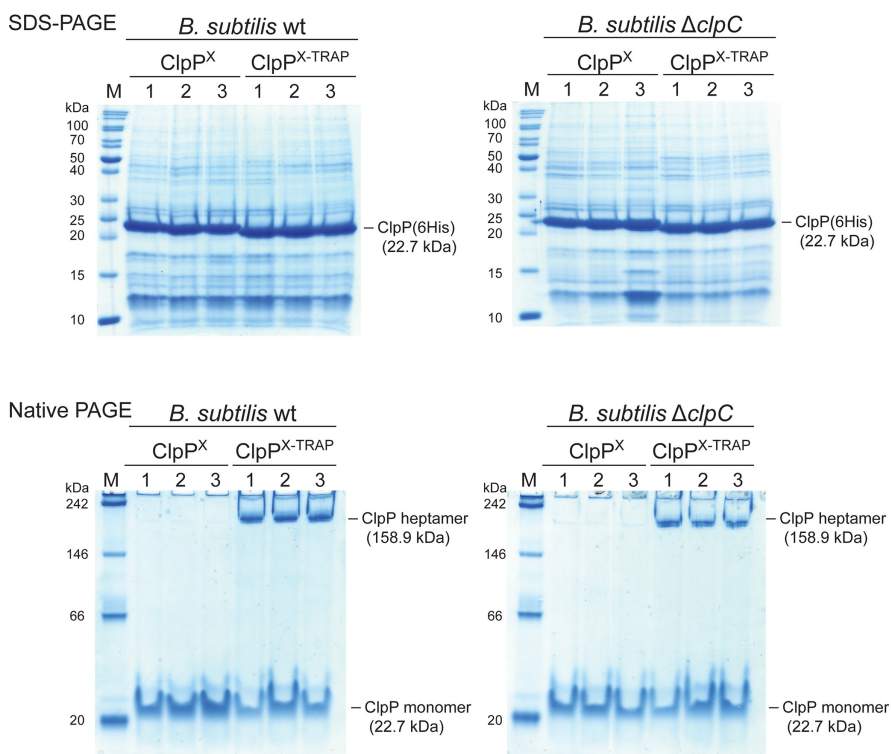
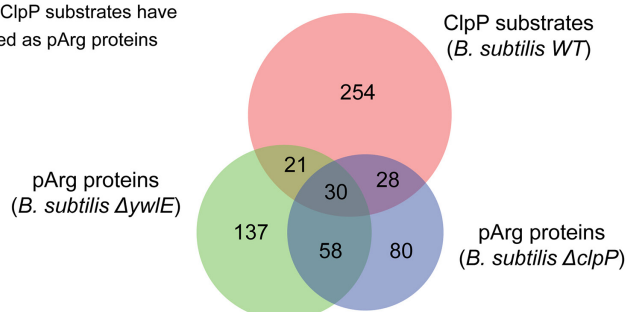
Data reporting. Source Data for Figs 1–5 are provided in the Supplementary Information. No statistical methods were used to predetermine sample size. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

35. Nguyen, H. D. *et al.* Construction of plasmid-based expression vectors for *Bacillus subtilis* exhibiting full structural stability. *Plasmid* **54**, 241–248 (2005).
36. Gerth, U. *et al.* Fine-tuning in regulation of Clp protein content in *Bacillus subtilis*. *J. Bacteriol.* **186**, 179–191 (2004).
37. Msadek, T. *et al.* ClpP of *Bacillus subtilis* is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. *Mol. Microbiol.* **27**, 899–914 (1998).
38. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
39. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925 (2007).
40. Taus, T. *et al.* Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* **10**, 5354–5362 (2011).
41. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
42. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
43. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).
44. Vizcaino, J. A. *et al.* The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069 (2013).
45. Suzuki, Y., Takeda, Y. & Ikuta, T. Immunoblotting conditions for human hemoglobin chains. *Anal. Biochem.* **378**, 218–220 (2008).
46. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
47. Nørby, J. G. Coupled assay of Na⁺/K⁺-ATPase activity. *Methods Enzymol.* **156**, 116–119 (1988).
48. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
49. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
50. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
51. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
53. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
54. Debreczeni, J. E. & Emsley, P. Handling ligands with Coot. *Acta Crystallogr. D* **68**, 425–430 (2012).
55. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D* **65**, 1074–1080 (2009).
56. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Schrodinger, LLC. *The PyMOL Molecular Graphics System* v.1.3r1 (2010).
58. Lee, B. G. *et al.* Structures of ClpP in complex with acyldepsipeptide antibiotics reveal its activation mechanism. *Nat. Struct. Mol. Biol.* **17**, 471–478 (2010).
59. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**, 779–790 (1993).
60. Rittinger, K. *et al.* Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Mol. Cell* **4**, 153–166 (1999).

a Expression of ClpP(6His)^{TRAP} in *B. subtilis***b Engineering of the ClpP cross mutant (ClpP^X)****c Expression of the ClpP^{X-TRAP} in *B. subtilis***

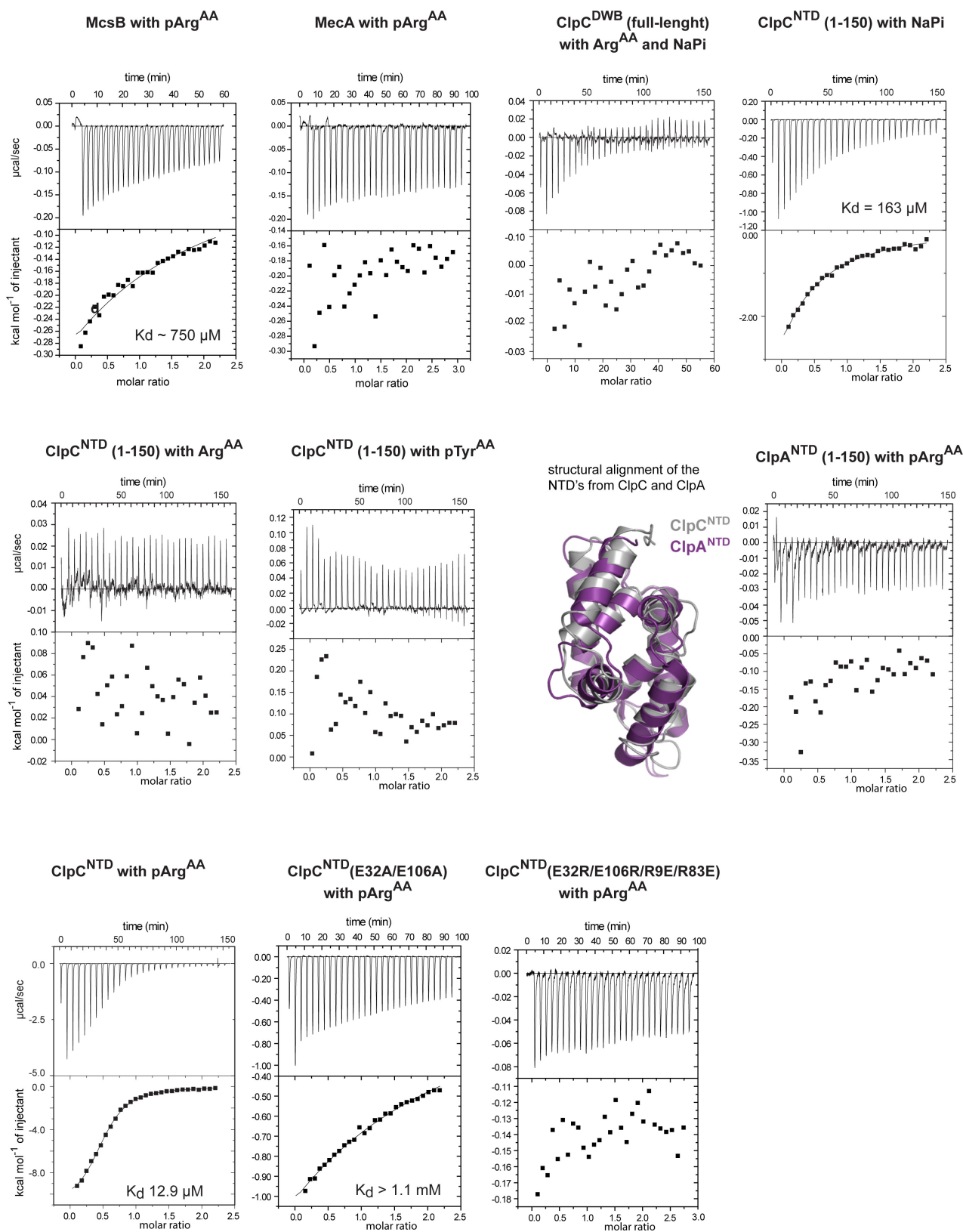
Extended Data Figure 1 | Development and validation of the ClpP^{X-TRAP} mutant. **a**, Co-NTA purification of wild-type and TRAP mutant ClpP(6His) expressed in heat-stressed *B. subtilis* wild-type cells. SDS-PAGE (left) reveals co-purification of two ClpP protein species. MALDI-MS analysis (bottom) estimates that the mass difference between the two ClpP species was 1,036 and 1,042 Da for wild-type and TRAP ClpP, respectively. These values fit to the expected mass of the 6His tag (1,065 Da), indicating that the purified proteins correspond to recombinant (tagged) and endogenous ClpP. Native-PAGE (right) shows that the two proteins are present predominantly as a composite, heptameric complex. Therefore, ClpP(6His)^{TRAP} expression in *B. subtilis* under heat-stress conditions is complicated by the formation of mixed complexes with endogenous (active) ClpP (cartoon). As the untagged endogenous ClpP was observed at similar levels as ClpP(6His), it is unlikely that efficient substrate-trapping complexes—built up exclusively by inactive ClpP^{TRAP}—are formed *in vivo*. **b**, Engineering of the ClpP cross mutant. Crystal structure of the ClpP heptamer (PDB code 3KTI; ref. 58) shown in top (left) and side (right) view. Zoomed-in picture shows interaction between Arg142 and Glu119 of two neighbouring subunits. To avoid the interaction of the recombinant ClpP(6His) trapping mutant with endogenous ClpP, these two residues were inter-exchanged (Glu119Arg/Arg142Glu),

thus leading to an electrostatic repulsion between the cross-mutant ClpP(6His)^{X-TRAP} and wild-type ClpP, while allowing formation of the respective homo-heptamers, as schematically indicated. **c**, Co-NTA purification of ClpP(6His)^X and ClpP(6His)^{X-TRAP} expressed in heat-stressed *B. subtilis* wild-type cells. SDS-PAGE (left) shows that the X variants do not co-purify with endogenous ClpP, demonstrating that the Glu119Arg/Arg142Glu mutation prevents the formation of hetero-oligomers (see cartoon). Native-PAGE analysis (middle) suggests that the ClpP(6His)^X protein has a reduced heptamerization propensity. However, the inactive version (Ser98Ala, “TRAP”) of the ClpP(6His)^X mutant is present predominantly as a heptameric complex, probably owing to the stabilization by trapped substrates. ClpP(6His)^{X-TRAP} thus represents a tool to trap substrates in the wild-type background of *B. subtilis*. Our experimental approach has the advantage of avoiding the use of the $\Delta clpP$ *B. subtilis* strain, which has an extremely pleiotropic phenotype (including increased levels of McsB and McsA¹⁴) that would largely bias the characterization of ClpP substrates. The ClpP^X protein represents the ideal negative control: it has reduced ability to heptamerize and therefore to degrade proteins, and its overexpression is expected to have little effect on the overall levels of protein degradation and consequently on the abundance of substrate proteins.

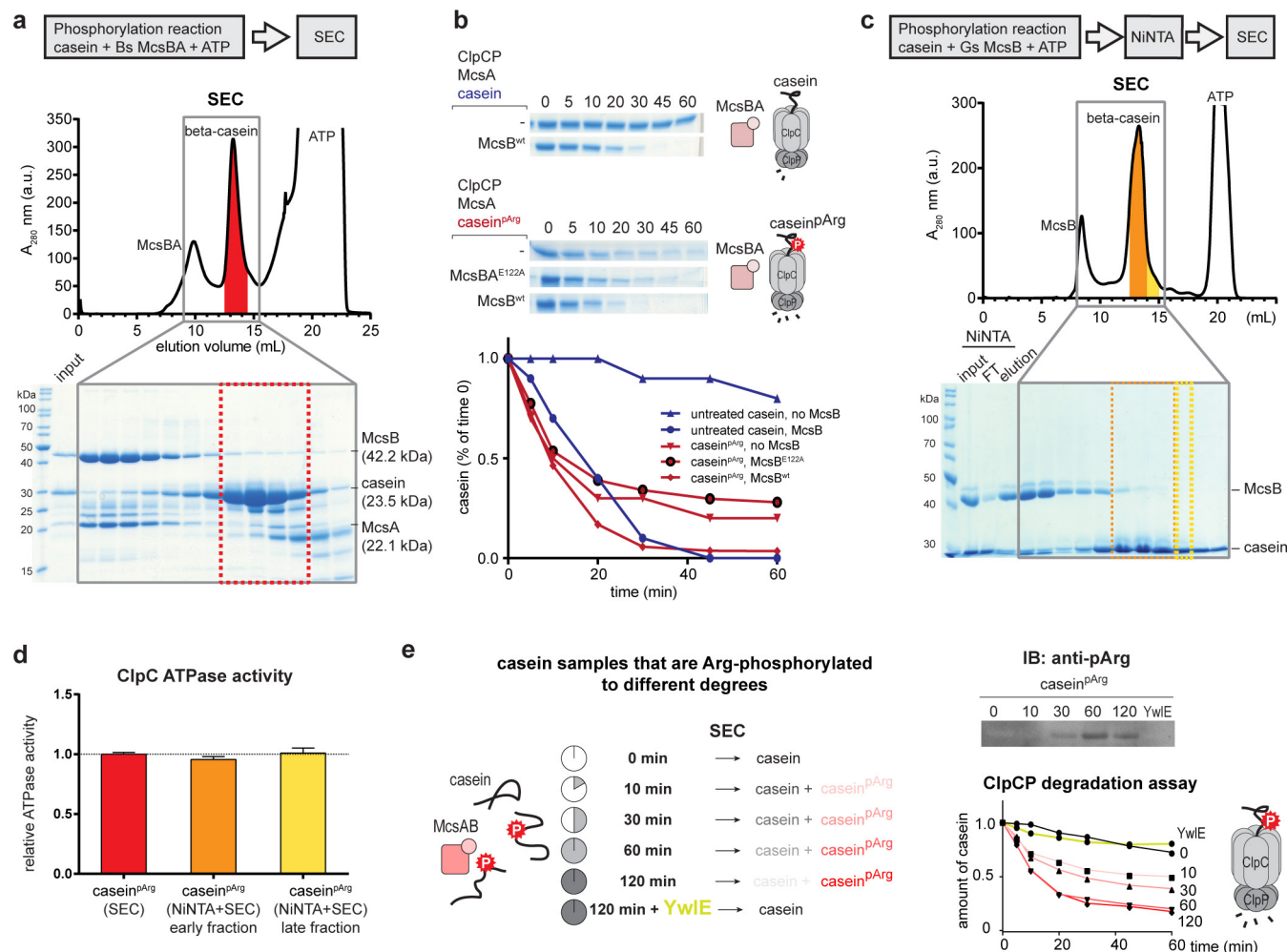
a Experiment 1: ClpP trapping mutant pull-downs in *B. subtilis* wt**b Experiment 2: ClpP trapping mutant pull-downs in *B. subtilis* wt vs. $\Delta clpC$** **c** 79 from 333 ClpP substrates have been identified as pArg proteins

Extended Data Figure 2 | *In vivo* characterization of ClpP^{X-TRAP} variants used in pull-down experiments. **a**, SDS-PAGE (left) and native-PAGE (right) analysis of co-NTA purifications of experiment 1: pull-down of His-tagged ClpP variants in wild-type *B. subtilis*. Numbers denote biological replicates. Each sample represents approximately 10% of the

total purification. **b**, SDS-PAGE (top) and native-PAGE (bottom) analysis of co-NTA purifications of experiment 2: pull-down of His-tagged ClpP variants in the wild-type (left) and $\Delta clpC$ (right) *B. subtilis* strain. Each sample represents approximately 5% of the total purification.

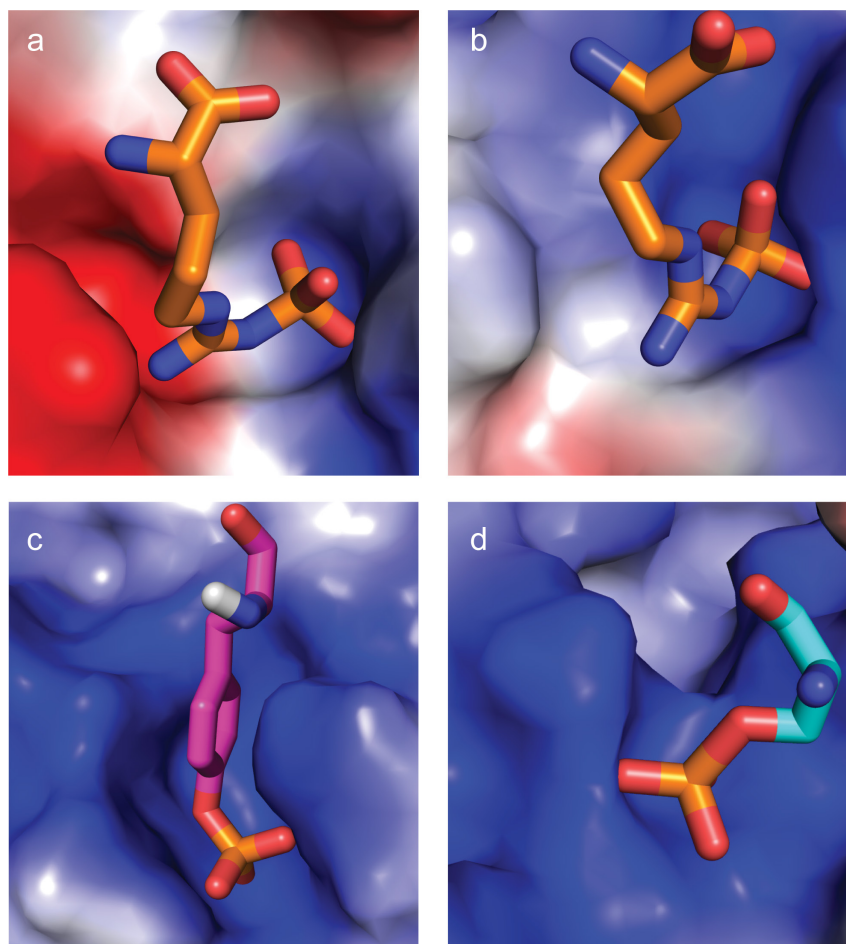


Extended Data Figure 3 | ITC binding data. For each binding study, the analysed interactions are indicated on the top, and the respective K_d values, when detected, are shown below. For reference, a structural alignment of the ClpC (grey) and ClpA (purple; PDB code 1K6K) NTDs, showing a high degree of structural similarity, is presented.



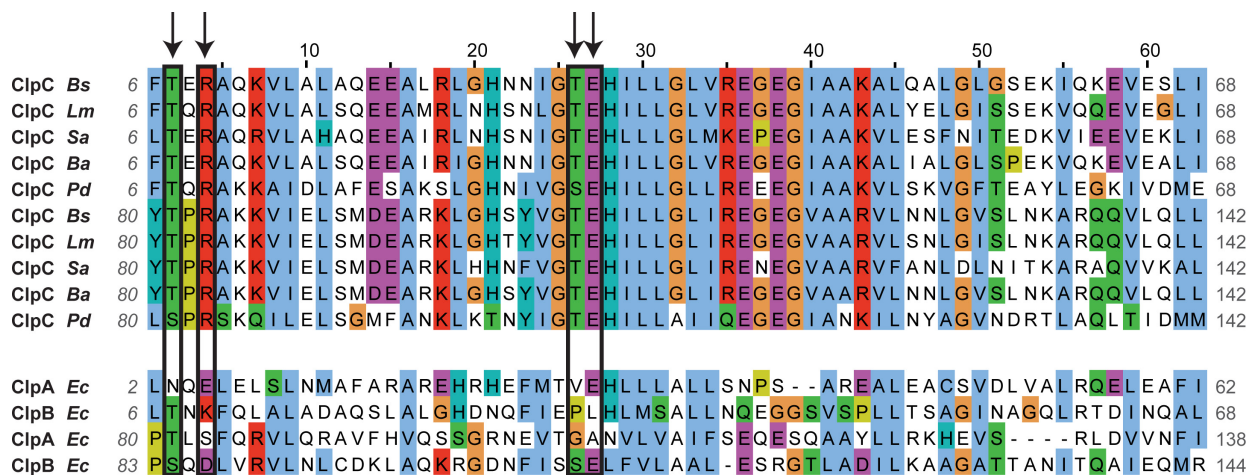
Extended Data Figure 4 | Preparation and validation of casein^{PArg} as a model substrate of ClpCP. **a**, Top, size exclusion chromatography (SEC) separation of casein^{PArg} from the *B. subtilis* McsBA complex after *in vitro* phosphorylation. The red markings indicate the fractions that were pooled. Bottom, SDS-PAGE analysis of the fractions indicates that the SEC procedure could separate at least 95% of the McsB protein from the pooled β -casein fractions. **b**, Evaluating the effect of the McsB contamination on the degradation of casein^{PArg}. Top, ClpCP *in vitro* degradation assay towards untreated β -casein. Middle, ClpCP *in vitro* degradation of casein^{PArg}, with and without additional McsB kinase. Bottom, addition of either inactive or active McsB did not have an effect on the initial degradation rate of casein^{PArg}. Of note, the degradation of untreated casein in the presence of McsBA is slightly delayed compared to the degradation of pre-modified casein^{PArg}. **c**, An alternative, improved protocol for purifying casein^{PArg}. After *in vitro* phosphorylation of β -casein by *G. stearothermophilus* McsB(6His), a Ni-NTA column was used to capture the tagged kinase. The flow-through fraction was then applied

to a SEC column to further separate remaining McsB from β -casein. Two different fractions of the β -casein peak were collected, the later-eluting one (yellow) having a higher degree of purity in relation to the earlier one (orange). **d**, Activation of the ClpC ATPase activity by the different casein^{PArg} preparations (2 μ M concentration each). Each fraction contains a different (substoichiometric) amount of the McsB contamination. ATPase measurements reveal an almost identical activation of ClpC for all fractions, indicating that the residual amounts of McsB present in the casein^{PArg} sample do not contribute to ClpC activation. Error bars denote standard deviation of technical triplicates. **e**, To obtain substrate samples varying in the amount of arginine phosphorylation, casein was pre-incubated with McsBA for increasing times. The YwIE arginine phosphatase was added to the sample phosphorylated most strongly (120 min incubation with McsBA) as a negative control. The resultant casein^{PArg} samples were subjected to pArg immunoblots using a pArg-specific antibody²⁴ (right, top) and to ClpCP degradation assays (right, bottom).



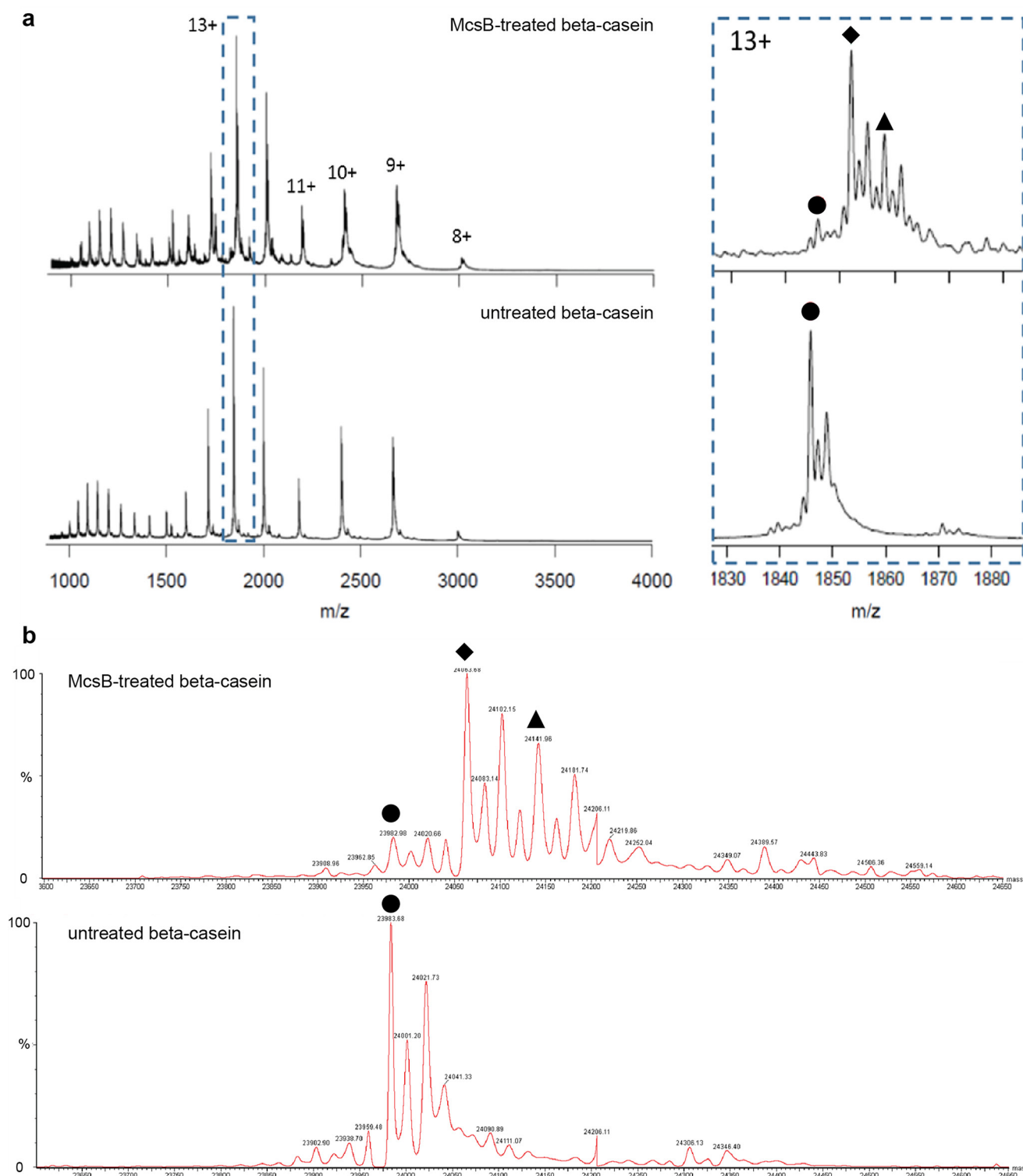
Extended Data Figure 5 | Binding pockets of pArg, pTyr and pSer/Thr. The binding pockets are shown as surface representation and coloured according to their electrostatics as calculated with PyMol (blue: positive, red: negative). Bound phosphoamino acids are presented as sticks with nitrogens and oxygens coloured blue and red, respectively. **a, b**, pArg-binding sites 1 and 2, respectively, of the ClpC^{NTD} domain. The sites are characterized by a 'bipolar' architecture with both a positive and a negative

area, jointly required to recognize a pArg side chain. **c, d**, pTyr-binding site of the Src SH2 domain (PDB code 1SPS; ref. 59) and pSer/Thr-binding site of the 14-3-3 domain (PDB code 1QJB; ref. 60). Both pTyr and pSer were part of a peptide but are shown in isolation for clarity. In contrast to the pArg-binding site, pTyr- and pSer/Thr-specific pockets are uniformly positively charged.



Extended Data Figure 6 | Sequence alignment of the pArg-binding site of ClpC from different species and of the homologous regions of other Clp ATPases. The two symmetrical regions (comprising residues 6–68 and 80–142, approximately) of each protein are aligned. Residues interacting with the pArg molecule (the Glu residue binding to the guanidinium group and the Arg/Thr residues interacting with the phosphate) are circled in black and marked by an arrow. Each of the two pArg-binding

sites comprises Glu and Thr from one symmetrical region and Arg and Thr from the other. The alignment shows high conservation of the critical residues of ClpC proteins from different MscB-containing bacteria (*B. subtilis*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Bacillus anthracis* and *Peptoclostridium difficile*). Conversely, the residues are not conserved in related Clp proteins (ClpA and ClpB) from MscB-deficient, Gram-negative bacteria.



Extended Data Figure 7 | Intact mass analysis of McsB-treated and untreated (control) β -casein. a, Unprocessed MS spectra.

Zoomed view of the 13+ charge state species shows two predominant arginine phosphorylation states in the McsB-treated sample (top): 1 phosphorylation (diamond, m/z 1,852) and 2 phosphorylations (triangle,

m/z 1,862), while only the non-pArg form (circle, m/z 1,848) can be visualized in the untreated control (bottom). **b**, Deconvoluted MS spectra, showing the average proportion of unmodified (circle, 23,982 Da), 1 pArg-containing (diamond, 24,063 Da) and 2 pArg-containing (triangle, 24,142 Da) β -casein.

Extended Data Table 1 | Summary of identified phospho-sites in various ClpP-trapping experiments

Uniprot	Gene Id	Protein Descriptions	Experiment 1		Experiment 2			
			<i>B. subtilis</i> wt		<i>B. subtilis</i> wt		<i>B. subtilis</i> Δ clpC	
			control	TRAP	control	TRAP	control	TRAP
Arg-phosphorylated proteins								
P71006	albF	Putative zinc protease AlbF		R17				
P37941	bfmBAB	2-oxoisovalerate dehydrogenase subunit b				R23		
O32222	csoR	Cu-sensing transcriptional repressor CsoR		R24		R24		
P37568	ctsR	Transcriptional regulator CtsR		R55		R55		R55
P42416	iolE	Inosose dehydratase		R216				
O34645	melA	Alpha-galactosidase		R56		R56		
O34338	mntB	Mn transport ATP-binding prot. MntB		R132, (R73)		R132		
P16263	odhB	2-oxoglutarate dehydrogenase E2				R300		
P12875	rplN	50S ribosomal prot. L14		R17		R17		
P42060	rplV	50S ribosomal prot. L22		R11				
P12879	rpsH	30S ribosomal prot. S8		R47, R72, R79		R47, R72		
P27206	srfAA	Surfactin synthase subunit 1		R226		R226		
P33166	tuf	Elongation factor Tu		R384				
O34909	yerA	Putative adenine deaminase YerA				R14		
O34949	ykoM	Uncharact. HTH-type transcript. regulator YkoM						R89
P42430	ykyB	Uncharacterized prot. YkyB				R26		R26
P54391	ypiF	Uncharacterized prot. YpiF		R4				
O07019	yvfU	Uncharact. transcript. regulatory protein YvfU				R116, R143		
Phosphorylated proteins, likely pArg								
P13800	degU	Transcriptional regulatory prot. DegU				(R184)		
P25499	hrcA	Heat-inducible Transcriptional repressor HrcA		(R27)		(R27)		
Ser/Thr/Tyr-phosphorylated proteins								
P25953	comGA	ComG operon prot. 1						T178
O31461	ilvE	Branched-chain aa transaminase 1		T249				
P42956	mltA	PTS system mannitol-specific EIICB		S365				S365
P55874	rpmI	50S ribosomal prot. L35				S47		S47
P21476	rpsS	30S ribosomal prot. S19				T77		
O34607	sdaAA	Probable L-serine dehydratase, α chain						T99
P54334	xkdO	Phage-like element PBSX prot. XkdO						S295
O34833	yceH	Uncharacterized prot. YceH				S296	S296	S296
O32131	yunB	Sporulation prot. YunB		T165				
P39808	yvyG	Uncharacterized prot. YvyG				Y49	Y49	Y49
P42294	ysiB	Uncharacterized prot. YsiB					S88	

Extended Data Table 2 | Data collection and refinement statistics

	ClpC NTD bound to pArg (PDB:5HBN)
Data collection	
Space group	P 65
Cell dimensions	
a, b, c (Å)	84.6, 84.6, 29.9
α , β , γ (°)	90, 90, 120
Resolution (Å)	42.3-1.60 (1.66-1.60) ^a
R _{merge}	0.062 (0.55)
R _{meas}	0.064 (0.58)
R _{pim}	0.015 (0.17)
I/σ(I)	29.62 (4.73)
CC _{1/2}	0.999 (0.95)
Completeness (%)	0.98 (0.83)
Redundancy	18.6 (11.0)
Refinement	
Resolution (Å)	42.3-1.60 (1.66-1.60)
No. reflections	16,041 (1,349)
R _{work} / R _{free}	0.145 (0.195)/ 0.166 (0.253)
No. atoms incl. hydrogens	2,503
Protein	2,347
Ligands and ions	116
Water	40
B factors	28.7
Protein	28.1
Ligand/ion	35.1
Water	36.3
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	0.97

^aValues in parentheses are for highest-resolution shell.

Extended Data Table 3 | Summary of protein expression and purification conditions

Protein	Expression	Buffer A	Buffer B
Bs ClpC	37°C, 3h	25 mM Tris pH 7.5, 300 mM NaCl	25 mM Tris pH 7.5, 300 mM NaCl
Bs ClpC ^{DWB}	37°C, 3h	25 mM Tris pH 8, 300 mM NaCl	25 mM Tris pH 7.5, 300 mM NaCl
Bs ClpC ^{NTD}	37°C, 2h	25 mM Tris pH 8.5, 150 mM NaCl, 1 mM TCEP	5 mM Tris pH 8.5, 50 mM NaCl, 1 mM TCEP
Ec ClpA ^{NTD}	37°C, 2h	25 mM Tris pH 8.5, 150 mM NaCl, 1 mM TCEP	5 mM Tris pH 8.5, 50 mM NaCl, 1 mM TCEP
Bs ClpP	37°C, 3h	25 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol	25 mM Tris pH 7.5, 150 mM NaCl, 10% glycerol
Bs McsA	20°C, 5h	25 mM Tris pH 7.5, 300 mM NaCl, TCEP 1 mM	25 mM Tris pH 7.5, 100 mM NaCl, TCEP 1 mM
Bs McsB	18°C, O/N	25 mM Tris pH 7.5, 300 mM KCl	25 mM Tris pH 7.5, 50 mM KCl
Bs McsB ^{R212A}	18°C, O/N	25 mM Tris pH 7.5, 300 mM KCl	25 mM Tris pH 7.5, 50 mM KCl
Gs McsB	37°C, 3h	25 mM Hepes pH 7.5, 300 mM KCl	25 mM Hepes pH 7.5, 100 mM KCl
Gs McsB ^{R122A}	18°C, O/N	50 mM Tris pH 7.5, 50 mM NaCl	50 mM Tris pH 7.5, 50 mM NaCl
Bs MecA	25°C, 5h	25 mM Hepes pH 8, 300 mM KCl	25 mM Hepes pH 8, 300 mM KCl
Gs YwIE	37°C, 3h	25 mM Tris pH 7.5, 300 mM NaCl	25 mM Tris pH 7.5, 100 mM NaCl
Gs YwIE ^{D118N}	37°C, 3h	25 mM Tris pH 7.5, 300 mM NaCl	25 mM Tris pH 7.5, 100 mM NaCl

Bs, *Bacillus subtilis*; Ec, *Escherichia coli*; Gs, *Geobacillus stearothermophilus*; O/N, overnight.

Break-induced telomere synthesis underlies alternative telomere maintenance

Robert L. Dilley¹, Priyanka Verma¹, Nam Woo Cho¹, Harrison D. Winters¹, Anne R. Wondisford¹ & Roger A. Greenberg^{1,2}

Homology-directed DNA repair is essential for genome maintenance through templated DNA synthesis. Alternative lengthening of telomeres (ALT) necessitates homology-directed DNA repair to maintain telomeres in about 10–15% of human cancers. How DNA damage induces assembly and execution of a DNA replication complex (break-induced replisome) at telomeres or elsewhere in the mammalian genome is poorly understood. Here we define break-induced telomere synthesis and demonstrate that it utilizes a specialized replisome, which underlies ALT telomere maintenance. DNA double-strand breaks enact nascent telomere synthesis by long-tract unidirectional replication. Proliferating cell nuclear antigen (PCNA) loading by replication factor C (RFC) acts as the initial sensor of telomere damage to establish predominance of DNA polymerase δ (Pol δ) through its POLD3 subunit. Break-induced telomere synthesis requires the RFC-PCNA-Pol δ axis, but is independent of other canonical replisome components, ATM and ATR, or the homologous recombination protein Rad51. Thus, the inception of telomere damage recognition by the break-induced replisome orchestrates homology-directed telomere maintenance.

Tremendous progress has been made in identifying the events responsible for recognizing and repairing DNA double-strand breaks (DSBs)¹. A complex aspect of this response is homology-directed DNA repair (HDR), which can involve numerous possibilities to capture homologous regions of the genome to use for templated DNA synthesis and repair. The detailed order of molecular events that ensues after the initial sensing of DSBs to allow the execution of homology-directed synthesis remains enigmatic. Specifically, how the DNA damage response coordinates productive interactions between DNA replication complexes to perform break-induced DNA synthesis has not been extensively demonstrated in mammalian cells. ALT is a clinically relevant example of a DNA repair pathway that requires homology-directed synthesis to maintain telomeres in ~10–15% of human cancers^{2,3}. Additionally, such synthesis could represent an attractive therapeutic target against cancers, especially if it proves to be different from canonical S-phase replication.

Telomere breaks stimulate long-tract synthesis

To study homology-directed synthesis at ALT telomeres, we developed a bromodeoxyuridine (BrdU) pulldown approach to isolate and quantify nascent telomeres synthesized following telomere-targeted DSBs generated by the fusion of the Shelterin component TRF1 to the FokI endonuclease (Fig. 1a). Using stable ALT-positive U2OS cell lines expressing TRF1-FokI under tetracycline-control, a 2-h damage induction with wild-type TRF1-FokI, but not the FokI(D450A) nuclease-null mutant, resulted in a ~10-fold increase in nascent telomere synthesis in asynchronous and G2-enriched cells (Fig. 1b, c and Extended Data Fig. 1a–g). Concurrent synthesis of nascent C- and G-rich telomere strands was evident from 1 h post-induction and became maximal at ~2 h (Fig. 1d). Nascent Alu repeat DNA was not increased by TRF1-FokI expression, demonstrating the specificity of break-induced telomere synthesis (Fig. 1d).

To understand the nature of individual DNA synthesis events, we adapted the single-molecule analysis of replicated DNA (SMARD) technique for studying break-induced telomere synthesis^{4,5}. After

induction of TRF1-FokI, U2OS cells were sequentially incubated with iododeoxyuridine (IdU) and chlorodeoxyuridine (CldU), genomic DNA was digested, and telomere fragments were isolated on the basis of size (Fig. 1e). The percentage of telomeres with IdU/CldU incorporation increased with the duration of TRF1-FokI induction (Fig. 1f, g). Break-induced telomere synthesis proceeded in a unidirectional fashion, often to the end of the telomere fragment. Nascent telomere tracts ranged in length from 5 to 70 kilobases (kb), with a median value of 19.8 kb ($n = 46$) that matched the median length of the overall telomere fibres observed (20.1 kb; $n = 45$) (Fig. 1h). Furthermore, ~80% of nascent telomere fragments were completely labelled and ~98% of nascent fragments had label on at least one of the ends. Taken together, these data suggest that DSBs at ALT telomeres induce long-tract telomeric DNA synthesis.

As a complementary approach, BrdU immunofluorescence at telomeres provides a means to assess spontaneous synthesis of ALT telomeres. Cell lines that utilize ALT, but not telomerase, displayed elevated BrdU incorporation at telomeres in a pattern distinct from S-phase replication (Extended Data Fig. 2a, b, d), consistent with previous reports⁶. TRF1-FokI expression increased analogue incorporation at ALT telomeres in interphase and metaphase cells (Extended Data Fig. 2c, e), suggesting that telomere damage may be an initiating event for spontaneous ALT telomere synthesis⁷.

Expanding on our observations, we generated a panel of TRF1-FokI inducible lines from cells that either utilize ALT (U2OS, VA13, SKNF1) or telomerase (HeLa 1.3, HeLa S3, 293T) for telomere maintenance. Notably, all lines showed evidence of break-induced telomere synthesis by BrdU pulldown upon induction with TRF1-FokI (Extended Data Fig. 3a–c). This holds true not only across telomere maintenance mechanism, but also regardless of ATRX status, overall telomere length differences, and cell type (Extended Data Fig. 3a). Notably, a recent study provided evidence that replication stress can activate ALT mechanisms in primary and telomerase-positive cells⁸. We propose that although any cell may have the capacity for break-induced telomere synthesis, ALT-positivity entails greater levels of telomere damage that promotes

¹Department of Cancer Biology, Abramson Family Cancer Research Institute, Bassett Research Center for BRCA, Perelman School of Medicine, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. ²Department of Pathology, Abramson Family Cancer Research Institute, Bassett Research Center for BRCA, Perelman School of Medicine, University of Pennsylvania, 421 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA.

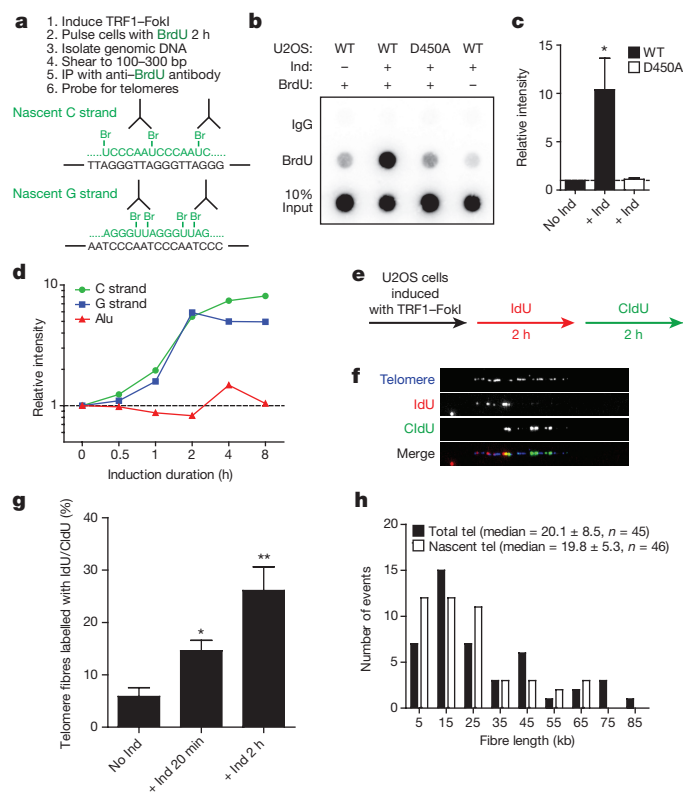


Figure 1 | Break-induced telomere synthesis occurs by long-tract unidirectional telomeric recombination. **a**, Schematic of BrdU pulldown. IP, immunoprecipitation. **b–d**, BrdU pulldown dot blot for telomere content using a 32 P-labelled telomere oligonucleotide (**b**) from U2OS cells induced (Ind) with TRF1-FokI for 2 h, with quantification (**c**) and time course of C- and G-rich telomere strands compared to Alu repeats (**d**). **e**, Schematic of telomere SMARD. **f–h**, Representative images (**f**) of telomere (blue) labelled with IdU (red) and CldU (green) from U2OS cells induced (Ind) with TRF1-FokI, with quantification (**g**) and length of telomere fibres (**h**). Median length quantified \pm 95% C.I. WT, wild-type; D450A, nuclease-null mutant. Data represent mean \pm s.e.m. of three (**c**) or two (**g**) independent experiments. $^{**}P \leq 0.01$, $^{*}P \leq 0.05$.

homology-directed DNA synthesis and telomere maintenance. Therefore, non S-phase telomere synthesis (Extended Data Fig. 2a, b, d) is apparent at baseline only in ALT cells.

Break-induced telomere synthesis by alternative HDR

Genetic studies in yeast demonstrated that break-induced replication is responsible for telomere recombination, which can proceed through Rad51-dependent and -independent mechanisms^{9–15}. Rad51, together with the Hop2–Mnd1 heterodimer, localize to ALT-associated PML bodies (APBs) and facilitate long-range telomere movement and clustering in ALT cells^{7,16,17}. Cells lacking Hop2 from CRISPR–Cas9-mediated excision showed reduced telomere clustering, APB formation, and telomere exchanges in ALT-positive VA13 cells (Extended Data Fig. 4a–f). ATR is a damage-sensing kinase that signals replication stress and is important for ALT telomere integrity and cell survival¹⁸. Disruption of ATR and Chk1 signalling by knockdown and small-molecule inhibitors reduced Hop2 recruitment to telomeres after TRF1–FokI induced damage, whereas ATM disruption had no effect (Fig. 2a). Similarly, ATR knockdown restricted telomere mobility after TRF1–FokI induction in U2OS cells (Fig. 2b), thus implicating ATR and Rad51–Hop2 as critical for ALT telomere mobility.

We next asked whether ATR and Rad51–Hop2 are required for break-induced telomere synthesis. Surprisingly, ATR, Rad51, and Hop2 were all dispensable for synthesis. Conversely, knockdown of each gene paradoxically increased levels of nascent telomeres, which held true over an 8-h time course (Fig. 2c, Extended Data Fig. 5a).

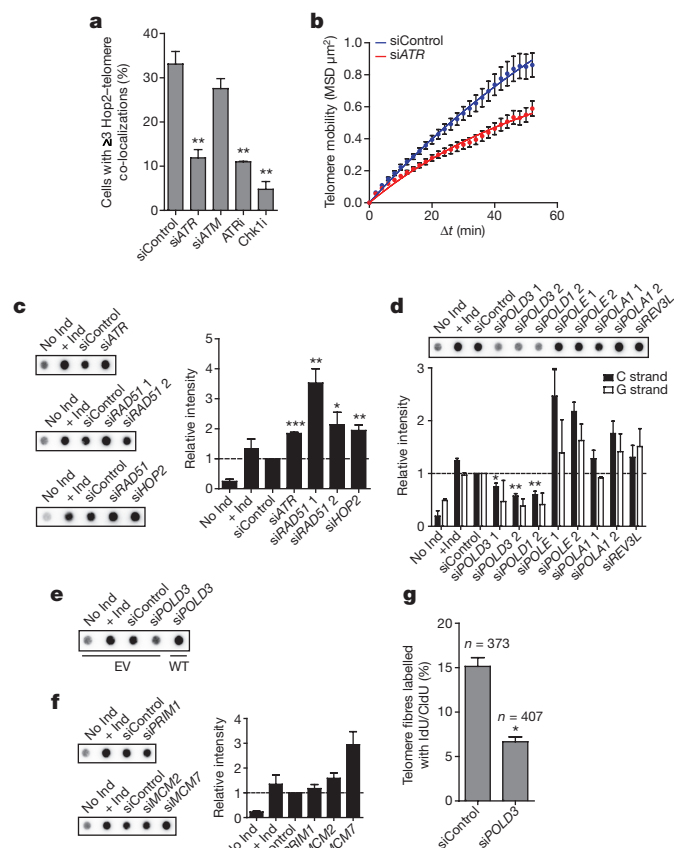


Figure 2 | Break-induced telomere synthesis occurs by alternative HDR and utilizes a non-canonical replisome defined by Pol δ .

a, Quantification of co-localized Hop2 and telomere foci in TRF1–FokI expressing VA13 cells treated with the indicated small interfering RNAs (siRNAs) or inhibitors. **b**, Mean squared displacement (MSD) analysis of live-cell telomere movement from U2OS cells following TRF1–FokI induction. **c–f**, BrdU pulldown dot blots for telomere content (**c**, **d**, **e**, **f**) from U2OS cells induced (Ind) with TRF1–FokI for 2 h, with rescue experiment (**e**), and quantifications (**c**, **d**, **f**). **g**, Quantification of telomere SMARD from U2OS cells induced with TRF1–FokI and treated with the indicated siRNAs. EV, empty vector; WT, reconstituted POLD3. Data represent mean \pm s.e.m. of at least two independent experiments. $^{***}P \leq 0.001$, $^{**}P \leq 0.01$, $^{*}P \leq 0.05$.

Similarly, spontaneous ALT telomere synthesis did not require Rad51 (Extended Data Fig. 5e, i). To investigate the long-term consequences of depletion of this pathway, we examined the telomere length of VA13 HOP2 CRISPR clones. All of the 6 clones lacked detectable Hop2 protein expression, with no telomere shortening observed at approximately population doubling (PD) 25 or longer time points (Extended Data Fig. 4g, h). Collectively, this provides evidence for Rad51-independent mechanisms of mammalian break-induced telomere synthesis and ALT telomere maintenance. Although ATR regulates damage signalling, telomere integrity, and survival in ALT cells, our data suggest it is not an essential component of the break-induced replisome at telomeres.

Break-induced telomere synthesis requires Pol δ

We next surveyed the replisome dependencies of break-induced telomere synthesis. Replicative DNA polymerases Pol δ , Pol ϵ , and Pol α -primase were previously implicated in yeast break-induced replication^{12,19}. Pol δ , including the POLD3 and POLD4 accessory and POLD1 catalytic subunits, was required for break-induced telomere synthesis (Fig. 2d, e, Extended Data Fig. 5a, b, d). Unexpectedly, Pol δ was required for synthesis of both C- and G-rich telomere strands, whereas Pol ϵ and Pol α -primase were dispensable as was the MCM2-7 replicative helicase (Fig. 2d, f). Notably, depletion of POLD3 resulted

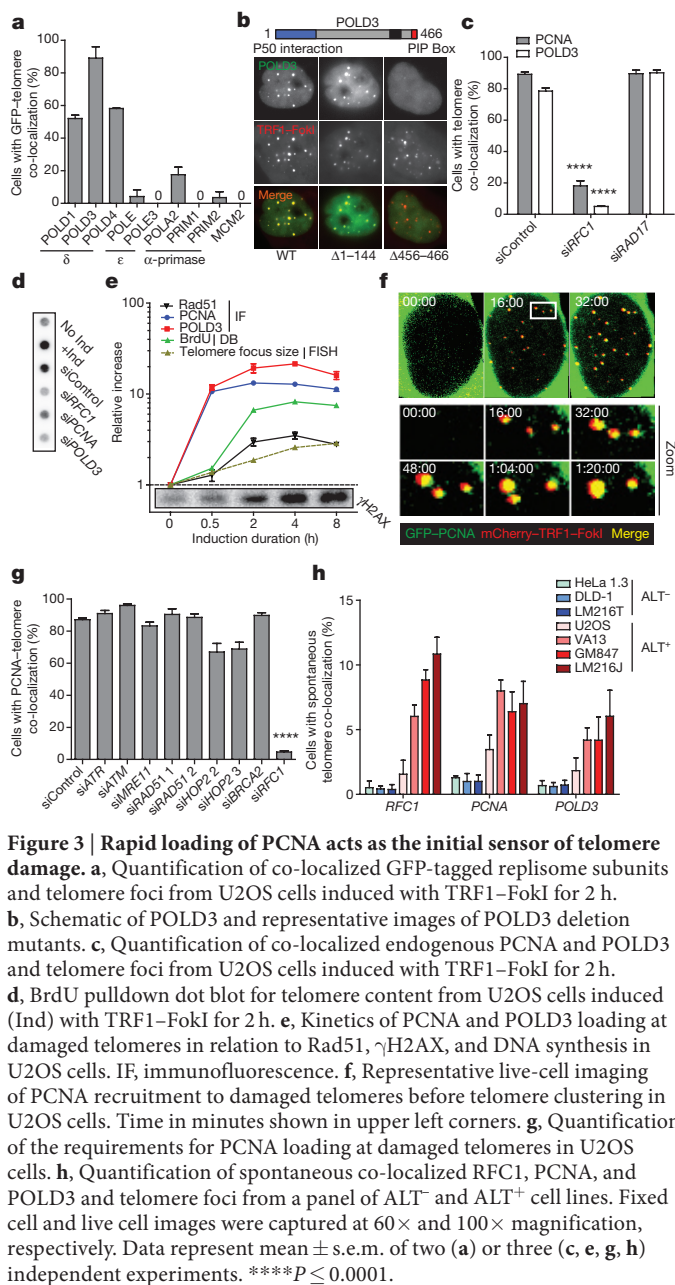


Figure 3 | Rapid loading of PCNA acts as the initial sensor of telomere damage. **a**, Quantification of co-localized GFP-tagged replisome subunits and telomere foci from U2OS cells induced with TRF1–FokI for 2 h. **b**, Schematic of POLD3 and representative images of POLD3 deletion mutants. **c**, Quantification of co-localized endogenous PCNA and POLD3 and telomere foci from U2OS cells induced with TRF1–FokI for 2 h. **d**, BrdU pulldown dot blot for telomere content from U2OS cells induced (Ind) with TRF1–FokI for 2 h. **e**, Kinetics of PCNA and POLD3 loading at damaged telomeres in relation to Rad51, γH2AX, and DNA synthesis in U2OS cells. IF, immunofluorescence. **f**, Representative live-cell imaging of PCNA recruitment to damaged telomeres before telomere clustering in U2OS cells. Time in minutes shown in upper left corners. **g**, Quantification of the requirements for PCNA loading at damaged telomeres in U2OS cells. **h**, Quantification of spontaneous co-localized RFC1, PCNA, and POLD3 and telomere foci from a panel of ALT[−] and ALT⁺ cell lines. Fixed cell and live cell images were captured at 60× and 100× magnification, respectively. Data represent mean ± s.e.m. of two (a) or three (c, e, g, h) independent experiments. *****P* ≤ 0.0001.

in ~2.5-fold less incorporation of IdU/CldU in telomere fibres after TRF1–FokI-induced breaks (Fig. 2g). POLD3 is also part of the Pol ζ complex involved in translesion synthesis^{20–22}. However, the catalytic subunit of Pol ζ (REV3L) as well as the other translesion synthesis proteins Pol η (POLH) and REV1 were not needed for break-induced telomere synthesis (Fig. 2d, Extended Data Fig. 5c). Therefore, the major function of POLD3 in break-induced telomere synthesis is through Pol δ. Notably, the requirements for break-induced telomere synthesis using TRF1–FokI faithfully recapitulate the requirements for spontaneous ALT telomere synthesis. Specifically, non S-phase telomere synthesis in three ALT lines required POLD3/Pol δ, but occurred independently of Pol ε, Pol α, and Pol ζ (Extended Data Fig. 5f–i). Collectively, these data define a non-canonical replisome involved in ALT telomere synthesis.

RFC–PCNA is the initial sensor of telomere damage

Pol δ showed robust recruitment to TRF1–FokI damage sites in U2OS cells, whereas Pol ε, Pol α-primase, and MCM2–7 were present at much lower levels (Fig. 3a, Extended Data Fig. 6a, b). POLD3 facilitates interaction of the Pol δ complex with the PCNA clamp for

processive synthesis and strand displacement²³. Notably, Pol δ has higher affinity for PCNA than does Pol ε, and PCNA is known to function in repair processes outside S-phase^{24,25}. Deletion of the PCNA-interacting peptide (PIP) box (Δ456–466) of POLD3 disrupted its recruitment to damage sites (Fig. 3b). Functionally, PCNA interaction with POLD3 facilitates recruitment of the whole Pol δ complex to damaged ALT telomeres (Extended Data Fig. 6c, d). The RFC1–5 clamp loading complex was required for PCNA–POLD3 telomere localization, whereas the alternative clamp loader subunit, Rad17 was dispensable (Fig. 3c). Furthermore, the entire axis consisting of RFC1–PCNA–POLD3 localized in an inducible fashion to ~90% of damaged ALT telomeres and was required for break-induced telomere synthesis (Fig. 3d, Extended Data Fig. 5h, 6e, f).

Both PCNA and POLD3 showed ~10-fold increases in telomere localization by 30 min after induction with TRF1–FokI (Fig. 3e). By contrast, Rad51 localization occurred more slowly and was maximal by 2 h after induction (Fig. 3e). Peak telomere synthesis coincided with an increase in DSB signalling (Figs 1d, 3e). Time-lapse imaging revealed that GFP–PCNA localized to TRF1–FokI damage sites soon after they became visible and before ALT telomere merging events (Fig. 3f, *n* = 20 cells). Consistent with PCNA loading being an early damage response, its localization was independent of proximal damage response factors ATR, ATM, MRE11, or homologous recombination proteins Rad51, Hop2, and BRCA2 (Fig. 3g). Importantly, RFC1, PCNA, and POLD3 spontaneously localized to ~2–10% of telomeres specifically in ALT-positive cells, consistent with the presence of persistent damage at a subset of ALT telomeres¹⁷ (Fig. 3h). These data reveal that PCNA loading is the initial damage sensor at ALT telomeres, thus establishing a platform to assemble the break-induced replisome.

POLD3 is critical for ALT telomere maintenance

Pol32, the yeast homologue of POLD3, is required for recombination dependent survivors of telomerase deficiency¹². Transient knockdown of POLD3 decreased spontaneous ALT telomere synthesis and single-telomere-exchange events by chromosome orientation fluorescence *in situ* hybridization (CO–FISH), but had no immediate effect on C-circles or telomere length (Extended Data Figs 5f–i, 7a, b). We investigated the consequences of prolonged POLD3 depletion on ALT telomere maintenance using CRISPR–Cas9 in U2OS cells. Although we were unable to generate surviving cells with complete loss of POLD3, we obtained 4 clones (c1–c4) with in-frame deletions and residual expression of POLD3 (Extended Data Fig. 7c–g). Notably, all 4 clones had reduced levels of the entire Pol δ complex (Fig. 4a, Extended Data Fig. 7g), consistent with a stabilizing role for POLD3 (ref. 26). Clones c1–c3 displayed accelerated telomere shortening at ~PD 25 compared to the empty guide control, whereas clone c4 had a more minor phenotype (Fig. 4b, Extended Data Fig. 7h, i). Telomere length in 5 clones with normal POLD3 expression (c5–c9) was unchanged at ~PD 25 (Extended Data Fig. 8a). The telomere shortening observed in clones c1–c3 is greater than expected for cells lacking a telomere maintenance mechanism, representing a loss of ~800–1200 bp of telomeric repeats per cell division. However, telomeres did not continue to shorten over time in these clones. This is consistent with accelerated shortening and stabilization observed in other ALT lines in which telomere maintenance mechanisms have been partially impaired²⁷. Additional U2OS POLD3 CRISPR clones from an independent guide RNA also displayed shortened telomeres compared to the parental line or clones derived from the same guide RNA that failed to exhibit reduced POLD3 expression, making it unlikely that the effects observed are due to clonal variation (Extended Data Fig. 8c). Collectively, analysis of the mean telomere length of the 31 clones from both of the guide RNAs revealed a significant decrease in clones with reduced POLD3 expression compared to those with normal POLD3 expression (Fig. 4c). In contrast, telomere length changes were not observed in 11 POLD3 CRISPR clones from telomerase-positive HeLa 1.3 cells (Extended Data Fig. 8d–f), suggesting

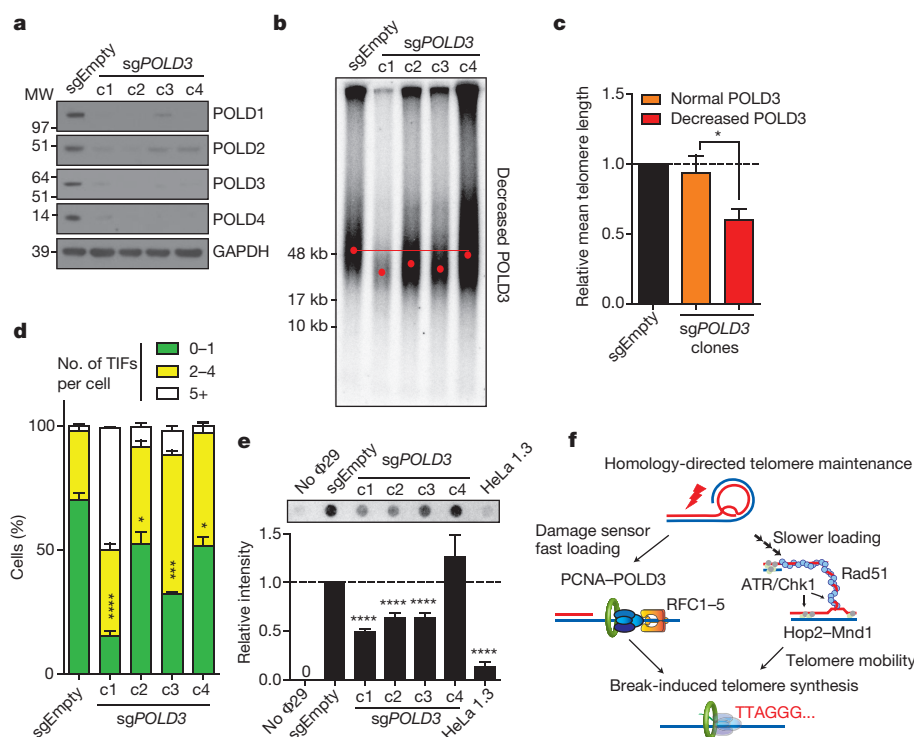


Figure 4 | POLD3 is critical for telomere maintenance in ALT-dependent cells. a, Analysis of Pol δ protein expression by western blot from U2OS CRISPR clones (sgPOLD3) with decreased POLD3 expression (c1–c4) compared to empty guide control (sgEmpty). GAPDH is used as a loading control. Darker exposure shown in Extended Data Fig. 7g. **b**, TRF analysis using pulsed-field gel electrophoresis from U2OS clones c1–c4 at ~PD 25. Peak intensity of telomere length is indicated by red dot.

c, Quantification of relative mean telomere length in 31 pooled sg POLD3 clones from two independent guide RNAs with normal and decreased POLD3 protein expression. **d**, **e**, Analysis of telomere maintenance by TIF quantification (co-localized 53BP1 and telomere foci) (**d**) and C-circle assay (**e**) from U2OS clones c1–c4. **f**, Model. Data represent mean \pm s.e.m. of three independent experiments. **** $P \leq 0.0001$, *** $P \leq 0.001$, * $P \leq 0.05$.

an increased requirement of POLD3 for processive telomere synthesis during ALT.

U2OS POLD3 CRISPR clones accumulated increased numbers of telomere dysfunction-induced foci (TIFs) (Fig. 4d, Extended Data Fig. 7i). C-circles are another marker of telomere maintenance specific to ALT-dependent cells²⁸. Clones c1–c3 had significantly decreased levels of C-circles that could be rescued by reconstituting wild-type POLD3, whereas clones c5–c9 did not display similar reductions (Fig. 4e, Extended Data Figs 7i–k, 8b). These data are consistent with a partial disruption of ALT activity and telomere maintenance in clones c1–c3 (Extended Data Fig. 7i). We propose that POLD3 is critical for the majority of nascent telomere synthesis during ALT and therefore underlies long-term telomere maintenance and ALT activity.

Discussion

Direct visualization of the dynamic process of ALT telomere recombination reveals that rapid RFC-mediated PCNA loading at damaged telomeres is the initial sensor of telomere damage, thus connecting DSB recognition with the assembly of a specialized replisome capable of executing break-induced telomere synthesis (Fig. 4f). On the other hand, Rad51 and Hop2 loading and long-range homology searches occur more slowly and are dispensable for the bulk of homology-directed DNA synthesis at telomeres (Fig. 4f). We postulate that PCNA can load at alternative structures with recessed 3'-ends and preferentially recruit Pol δ owing to its higher affinity for PCNA compared with that of Pol ϵ ²⁴. In contrast to yeast break-induced replication, Rad51 and Pol α -primase were not required for break-induced telomere synthesis. A repertoire of Rad51-independent mechanisms available to damaged telomeres, such as intra-telomere annealing or association with extra-chromosomal telomere repeats that are abundant in ALT cells may bypass the need for Rad51 and Pol α -mediated priming³. Persistent damage at ALT telomeres probably promotes Rad51 and other competing

repair mechanisms with differing kinetics of homology-directed telomere synthesis^{7,17} (Fig. 4f). We speculate that related processes are invoked at other vulnerable regions of the genome^{29–33}. The unique characteristics that differentiate this mechanism from scheduled S-phase replication may facilitate a better understanding of how alternative repair mechanisms enable genome evolution and enhance cancer cell fitness.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 May; accepted 16 September 2016.

Published online 19 October 2016.

- Ciccia, A. & Elledge, S. J. The DNA damage response: making it safe to play with knives. *Mol. Cell* **40**, 179–204 (2010).
- Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S. & Reddel, R. R. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* **14**, 4240–4248 (1995).
- Cesare, A. J. & Reddel, R. R. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat. Rev. Genet.* **11**, 319–330 (2010).
- Norio, P. & Schildkraut, C. Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science* **294**, 2361–2365 (2001).
- Sfeir, A. et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* **138**, 90–103 (2009).
- Nabetani, A., Yokoyama, O. & Ishikawa, F. Localization of hRad9, hHus1, hRad1, and hRad17 and caffeine-sensitive DNA replication at the alternative lengthening of telomeres-associated promyelocytic leukemia body. *J. Biol. Chem.* **279**, 25849–25857 (2004).
- Cho, N. W., Dilley, R. L., Lampson, M. A. & Greenberg, R. A. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* **159**, 108–121 (2014).
- O'Sullivan, R. J. et al. Rapid induction of alternative lengthening of telomeres by depletion of the histone chaperone ASF1. *Nat. Struct. Mol. Biol.* **21**, 167–174 (2014).
- Lundblad, V. & Blackburn, E. H. An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* **73**, 347–360 (1993).

10. Teng, S. C. & Zakian, V. A. Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**, 8083–8093 (1999).
11. Le, S., Moore, J. K., Haber, J. E. & Greider, C. W. *RAD50* and *RAD51* define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics* **152**, 143–152 (1999).
12. Lydeard, J. R., Jain, S., Yamaguchi, M. & Haber, J. E. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **448**, 820–823 (2007).
13. Malkova, A., Ivanov, E. L. & Haber, J. E. Double-strand break repair in the absence of *RAD51* in yeast: a possible role for break-induced DNA replication. *Proc. Natl Acad. Sci. USA* **93**, 7131–7136 (1996).
14. Chen, Q., Ijima, A. & Greider, C. W. Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. *Mol. Cell. Biol.* **21**, 1819–1827 (2001).
15. Teng, S. C., Chang, J., McCowan, B. & Zakian, V. A. Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Mol. Cell* **6**, 947–952 (2000).
16. Cox, K. E., Maréchal, A. & Flynn, R. L. SMARCA11 resolves replication stress at ALT telomeres. *Cell Reports* **14**, 1032–1040 (2016).
17. Yeager, T. R. *et al.* Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukaemia (PML) body. *Cancer Res.* **59**, 4175–4179 (1999).
18. Flynn, R. L. *et al.* Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science* **347**, 273–277 (2015).
19. Lydeard, J. R. *et al.* Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. *Genes Dev.* **24**, 1133–1144 (2010).
20. Johnson, R. E., Prakash, L. & Prakash, S. Pol31 and Pol32 subunits of yeast DNA polymerase δ are also essential subunits of DNA polymerase ζ . *Proc. Natl Acad. Sci. USA* **109**, 12455–12460 (2012).
21. Baranovskiy, A. G. *et al.* DNA polymerase δ and ζ switch by sharing accessory subunits of DNA polymerase δ . *J. Biol. Chem.* **287**, 17281–17287 (2012).
22. Makarova, A. V., Stodola, J. L. & Burgers, P. M. A four-subunit DNA polymerase ζ complex containing Pol δ accessory subunits is essential for PCNA-mediated mutagenesis. *Nucleic Acids Res.* **40**, 11618–11626 (2012).
23. Ducoux, M. *et al.* Mediation of proliferating cell nuclear antigen (PCNA)-dependent DNA replication through a conserved p21^{Cip1}-like PCNA-binding motif present in the third subunit of human DNA polymerase δ . *J. Biol. Chem.* **276**, 49258–49266 (2001).
24. Chilkova, O. *et al.* The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.* **35**, 6588–6597 (2007).
25. Karras, G. I. & Jentsch, S. The *RAD6* DNA damage tolerance pathway operates uncoupled from the replication fork and is functional beyond S phase. *Cell* **141**, 255–267 (2010).
26. Murga, M. *et al.* POLD3 is haploinsufficient for DNA replication in mice. *Mol. Cell* **63**, 877–883 (2016).
27. Zhong, Z.-H. *et al.* Disruption of telomere maintenance by depletion of the MRE11/*RAD50*/*NBS1* complex in cells that use alternative lengthening of telomeres. *J. Biol. Chem.* **282**, 29314–29322 (2007).
28. Henson, J. D. *et al.* DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. *Nat. Biotechnol.* **27**, 1181–1185 (2009).
29. Costantino, L. *et al.* Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* **343**, 88–91 (2014).
30. Mayle, R. *et al.* Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science* **349**, 742–747 (2015).
31. Minocherhomji, S. *et al.* Replication stress activates DNA repair synthesis in mitosis. *Nature* **528**, 286–290 (2015).
32. Miyabe, I. *et al.* Polymerase δ replicates both strands after homologous recombination-dependent fork restart. *Nat. Struct. Mol. Biol.* **22**, 932–938 (2015).
33. Verma, P. & Greenberg, R. A. Noncanonical views of homology-directed DNA repair. *Genes Dev.* **30**, 1138–1154 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Sfeir and A. Phillips (NYU) for guidance on telomere SMARD experiments and members of the Greenberg laboratory for critical discussion. This work was supported by NIH grants GM101149, CA138835, and CA17494 to R.A.G., who is also supported by funds from the Abramson Family Cancer Research Institute and Bassett Research Center for BRCA. R.L.D. was supported by NIH grants T32GM007170 and T32GM008216.

Author Contributions R.L.D., P.V., N.W.C., and R.A.G. designed the study. R.L.D. performed most of the experiments, with assistance from H.D.W. and A.R.W. P.V. conducted SMARD experiments. N.W.C. conducted ATR and Hop2 experiments. R.L.D., P.V., and R.A.G. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.G. (rogergr@mail.med.upenn.edu).

METHODS

Data reporting. No statistical methods were used to predetermine sample size.

Cell culture. U2OS, HeLa 1.3, HeLa S3, DLD-1, and 293T cell lines were grown in DMEM (Thermo Fisher) with 10% calf serum and 1% penicillin/streptomycin. VA13, GM847, LM216T, and LM216J cell lines were grown in DMEM (Thermo Fisher) with 10% FBS and 1% penicillin/streptomycin. SKNFI cell line was grown in RPMI (Thermo Fisher) with 10% FBS and 1% penicillin/streptomycin. VA13 cell line refers to WI-38 VA-13 subline 2RA. LM216T/J are matched lines. Cell lines were obtained from ATCC and tested negative for Mycoplasma using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza). The U2OS TRF1–FokI inducible cell line was authenticated by STR analysis (ATCC). Other lines were validated by ALT characteristics. None of the cell lines used is listed as commonly misidentified by the International Cell Line Authentication Committee (ICLAC).

ALT-positive lines used: U2OS, VA13, GM847, LM216J, SKNFI

ALT-negative lines used: HeLa 1.3 (long telomere), HeLa S3, DLD-1, LM216T, 293T

BrdU immunofluorescence. Cells were pulsed with 100 μ M BrdU (Sigma) for 2 h before fixation. After permeabilization, cells were denatured with 500 U ml⁻¹ DNaseI (Roche) in 1 \times reaction buffer (20 mM Tris-HCl (pH 8.4), 2 mM MgCl₂, 50 mM KCl in PBST) for 10–25 min at 37 °C in a humidified chamber. Coverslips were then washed and incubated with anti-BrdU antibody (BD) for 20 min at 37 °C followed by secondary antibody and telomere FISH. For metaphases, cells pulsed with BrdU were treated with 100 ng ml⁻¹ colcemid for 90 min followed by 75 mM KCl for 30 min. Cells were fixed in 3:1 methanol:acetic acid, dropped onto slides, and allowed to dry overnight. Denaturation was performed with 2N HCl for 30 min at room temperature followed by antibody incubations as described above.

BrdU pulldown dot blot. BrdU pulldown was adapted from a published protocol³⁴. Cells were pulsed with 100 μ M BrdU (Sigma) for 2 h before collection. Genomic DNA (gDNA) was isolated using phenol–chloroform extraction followed by resuspension in TE buffer. gDNA was then sheared into 100–300 bp fragments using a Covaris S220 sonicator. 1–4 μ g sheared gDNA was denatured for 10 min at 95 °C and cooled in an ice-water bath. Denatured gDNA was incubated with 2 μ g anti-IgG (Sigma) or anti-BrdU antibody (BD) diluted in immunoprecipitation buffer (0.0625% (v/v) Triton X-100 in PBS) rotating overnight at 4 °C. The next day, samples were incubated with 30 μ l Protein G magnetic beads (Pierce) that had been pre-bound to a bridging antibody (Active Motif) for 1 h rotating at 4 °C. Beads were subsequently washed three times with immunoprecipitation buffer and once with TE buffer. Beads were then incubated twice in elution buffer (1% (w/v) SDS in TE) for 15 min at 65 °C. Pooled eluate was cleaned with ChIP DNA Clean & Concentrator kit (Zymo). Samples, along with 10% inputs, were diluted into 2 \times SSC buffer, treated at 95 °C for 5 min, and dot-blotted onto an Amersham Hybond-N⁺ nylon membrane (GE). The membrane was then denatured in a 0.5 N NaOH 1.5 M NaCl solution, neutralized, and ultraviolet crosslinked. The membrane was hybridized with ³²P-labelled (TTAGGG)₆ oligonucleotides, unless otherwise noted, in PerfectHyb Plus Hybridization Buffer (Sigma) overnight at 37 °C. The next day, the membrane was washed twice in 2 \times SSC buffer, exposed onto a storage phosphor screen (GE Healthcare) and scanned using STORM 860 with ImageQuant (Molecular Dynamics). All quantifications were performed in Fiji and normalized to 10% input.

Telomere single-molecule analysis of replicated DNA (SMARD). The SMARD assay was performed as previously described^{4,5}. U2OS cells were induced with TRF1–FokI for 20 min or 2 h and were subsequently labelled by incubating with 30 μ M IdU for 2 h, followed by 30 μ M CldU for the next 2 h. After pulsing, 10⁶ labelled cells per condition were embedded in 1% agarose and lysed using detergents (100 mM EDTA, 0.2% sodium deoxycholate, 1% sodium lauryl sarcosine and 0.2 mg ml⁻¹ Proteinase K). The plugs were then washed several times with TE, treated with 100 μ M PMSE, and then washed again with TE buffer followed by incubation with 1 \times Cut-Smart buffer (NEB) for 30 min. The DNA in the plugs was then digested overnight at 37 °C using 50 U of both MboI and AluI (NEB) per plug. The digested plugs were then cast into a 0.7% low-melting point agarose gel and a distinct fragment running above 10 kb (containing telomeric DNA defined by Southern blotting) was excised, melted and stretched on slides coated with 3-aminopropyltriethoxysilane (Sigma-Aldrich). After denaturation of the DNA strands using alkali buffer (0.1 M NaOH in 70% ethanol and 0.1% β -mercaptoethanol), the DNA was fixed using 0.5% glutaraldehyde and incubated overnight with biotin-OO-(CCCTAA)₄ locked nucleic acid (LNA) probe (Exiqon) at 37 °C. Telomere FISH probes were then detected using the Alexa Fluor 405-conjugated streptavidin (Thermo-Fisher) followed by sequential incubation with the biotinylated anti-avidin antibody (Vector Laboratories) and additional Alexa 405-conjugated streptavidin. IdU and CldU were visualized using mouse anti-IdU (BD) and rat anti-CldU (Serotec) monoclonal antibodies followed by Alexa Fluor 568-goat anti-mouse and Alexa Fluor 488-goat anti-rat secondary antibodies (Life Technologies). Images were acquired using the NIS-element software (Nikon) and a Nikon eclipse

80i microscope equipped with a 63 \times objective and a Cool Snap camera (MYO). For calculating the length of the telomeres and replication tracts, the line-scan function from Image J was used. For conversion of microns to kilobases, as 10 bp (equals one turn of the helix) has a linear length of 3.4 nm, 0.26 microns corresponded to 1 kb of DNA.

Plasmids, primers, siRNAs, and CRISPR sgRNAs. Death domain (DD)–Oestrogen receptor (ER)–mCherry–TRF1–FokI and Flag–TRF1–FokI constructs were cloned as previously described⁷. Doxycycline-inducible TRF1–FokI lines were generated using the Tet-On 3G system. Briefly, Flag–DD–ER–mCherry–TRF1–FokI was cloned into the pLenti CMV TRE3G Puro Dest vector, which was introduced into cells engineered to co-express the reverse tetracycline transactivator 3G (rtTA3G). N-terminal GFP-tagged proteins were generated by PCR amplification and ligation of cDNAs from the ProQuest HeLa cDNA Library (Invitrogen) into the pDEST53 (Invitrogen) mammalian expression vector. CRISPR lines were generated using a two-vector system (pLentiCas9-Blast and pLentiGuide-Puro). POLD3 reconstitution vector was generated by cloning POLD3 cDNA (RefSeq NM_006591.2) into the pOZ–N–Flag–HA retroviral vector followed by site-directed mutagenesis of siRNA binding sites. Sanger sequencing of POLD3 CRISPR clones was performed on gDNA fragments cloned into a TOPO TA vector (Thermo Fisher).

Transient plasmid transfections were carried out with LipoD293 (Signagen), and siRNA transfections with Lipofectamine RNAiMax (Invitrogen) according to manufacturer's instructions. Analyses were performed 16 h after transfection of plasmids, and 72 h after siRNA transfection. All siRNAs were used at a final concentration of 20 nM.

The following primers were used for qRT-PCR:

POLD3 primer set 1: 5'-GAGTTCGTACGACGACCAAAAC-3', 5'-GCCA GACACCAAGTAGGTAAC-3';

POLD3 primer set 2: 5'-ACCAACAAGGAAACGAAACAGA-3', 5'-GG TTCCGTGACAGACACTGTA-3';

The following siRNA sequences were used:

Control siRNA (siCtrl): QIAGEN AllStars Negative Control siRNA;

ATR siRNA (siATR): 5'-AACCUCCGUGAUGUUGCUUGAdTdT-3';

ATM siRNA (siATM): 5'-GCGCCUGAUUCGAGAUCCUdTdT-3';

RAD51 siRNA (siRAD51): #1, 5'-UGUAGCAUAGCUCGAGCG-3', #2, 5'-CCA GAUCUGUCAUACGCUA-3';

HOP2 siRNA (siHOP2): #2, 5'-AAGAGAAGAUGUACGGCAA-3', #3, 5'-UCU GCUUAAAGGUGAAAGUAGCAGG-3';

BRCA2 siRNA (siBRCA2): 5'-GAAGAAUGCAGGUUUAAU-3';

RFC1 siRNA (siRFC1): 5'-GAAGGCGGCCUCUAAUACAUU-3';

RAD17 siRNA (siRAD17): 5'-CAGACUGGUUGACCAUUCU-3';

PCNA siRNA (siPCNA): 5'-GGAGGAAGCUGUUACCAUAAU-3';

MRE11 siRNA (siMRE11): Dharmacon SMARTpool M-009271-01-0005;

POLD3 siRNA (siPOLD3): #1, Invitrogen 4390824-s21045, #2, Invitrogen 4392420-s21046;

POLD1 siRNA (siPOLD1): #1, Invitrogen 4392420-s615, #2, Invitrogen 4392420-s616;

POLD4 siRNA (siPOLD4): 5'-GCAUCUCUAUCCCUAUGAUU-3';

POLE siRNA (siPOLE): #1, 5'-GGACAGGCGUACGAGUUCU-3'; #2, 5'-CU CGGAAGCUGGAAGAUUUU-3';

POLA1 siRNA (siPOLA1): #1, Invitrogen 4392420-s10772, #2, Invitrogen 4392420-s10774;

REV3L siRNA (siREV3L): 5'-CCCACUGGAAUUAUAGCACAAU-3';

PRIM1 siRNA (siPRIM1): Invitrogen HSS108448;

MCM2 siRNA (siMCM2): Invitrogen HSS106390;

MCM7 siRNA (siMCM7): Invitrogen HSS106405;

POLH siRNA (siPOLH): 5'-CTGGTTGTGAGCATTCCTGTGA-3';

REV1 siRNA (siREV1): 5'-ATCGGTGGAATCGGTTTGGA-3';

Knockdown efficiencies were evaluated by western blot (Extended data Fig. 9).

The following CRISPR sgRNA sequences were used:

sgPOLD3: #1, 5'-GCAGATAAAGCTGGTCCGCCA-3', #2, 5'-GAAATA TAGACGAGTTCGTCA-3';

sgHOP2: #1, 5'-GCCGAGCTGTAGTTGCTCG-3', #2, 5'-GCGGGAAA GGCGATGAGTAA-3', #3, 5'-GCGGGAGGTAACGGCGCGCT-3', #4, 5'-GAGT AGATTCACCCGTTGTC-3', #5, 5'-GACCATGAGAGCCGACAAC-3'.

Antibodies. The following antibodies were used: anti-BrdU (mouse B44, BD 347580; rat BU1/75, AbD Serotec OBT0030G), anti-ATRX (rabbit H-300, Santa Cruz sc-15408), anti-53BP1 (rabbit, Novus NB100-904), anti- γ H2AX (mouse JBW301, Millipore 05-636), anti-Flag (mouse M2, Sigma F1804), anti-PML (mouse PG-M3, Santa Cruz sc-966), anti-Rad51 (rabbit H-92, Santa Cruz sc-8349; mouse 14B4, Abcam ab-213), anti-Hop2/PSMC3IP (rabbit, Novus NBP1-92301), anti-POLD3 (mouse 3E2, Abnova H00010714-M01), anti-POLD1 (mouse 607, Abcam ab10362; rabbit, Bethyl A304-005A), anti-POLD2 (rabbit, Bethyl

A304-322A), anti-POLD4 (mouse 2B11, Abnova H00057804-M01A), anti-POLE (mouse 93H3A, Pierce MA5-13616; rabbit, Novus NBP1-68470), anti-POLE3 (rabbit, Bethyl A301-245A), anti-POLA1 (rabbit, Bethyl A302-851A), anti-MCM7 (rabbit, Bethyl A302-584A), anti-MCM4 (rabbit, Bethyl A300-193A), anti-MCM5 (rabbit, Abcam ab75975), anti-RFC1 (rabbit, Bethyl A300-320A), anti-PCNA (mouse PC10, CST #2586) anti-ATR (goat N-17, Santa Cruz sc-1887), anti-PRIM1 (rabbit H300, Santa Cruz sc-366482), anti-Rad17 (goat, Bethyl A300-151A), anti-REV3L (rabbit, GeneTex GTX100153), anti-POLH (rabbit, Bethyl A301-231A), anti-REV1 (rabbit H300, Santa Cruz sc-48806) anti-GAPDH (rabbit 14c10, CST #2118), anti- α Tubulin (mouse TU-02, Santa Cruz sc-8035).

Drugs. Doxycycline was used at a concentration of 40 ng ml⁻¹ for 16–24 h to induce expression of TRF1–FokI. Shield-1 (Cheminpharma LLC) and 4-hydroxytamoxifen (4-OHT) (Sigma-Aldrich) were both used at a concentration of 1 μ M for 2 h, unless otherwise stated, in to allow for TRF1–FokI stabilization and translocation into the nucleus. RO-3306 (Selleck Chemicals) was used at a concentration of 10 μ M for 20–24 h. G2 enrichment was confirmed by propidium iodide staining and flow cytometry. Colcemid (Roche) was used at a concentration of 100 ng ml⁻¹. The ATR inhibitor VE-821 (Selleck Chemicals) and Chk1 inhibitor LY2603618 (Selleck Chemicals) were used at a concentration of 5 μ M and 1 μ M respectively for 24 h.

Western blot. Cells were lysed in RIPA buffer supplemented with cOmplete protein inhibitor cocktail (Roche) and Halt phosphatase inhibitor cocktail (Thermo) on ice and subsequently spun down at max speed at 4 °C. The supernatant was removed and protein concentration determined using the Protein Assay Dye Reagent (Bio-Rad). 20–40 μ g of protein was run on a 4–12% Bis–Tris gel (Invitrogen). Proteins were transferred onto an Amersham Protran 0.2 μ m nitrocellulose membrane (GE) and blocked with 5% milk. Membranes were incubated with primary antibodies overnight at 4 °C. The next day membranes were incubated with secondary antibodies for 1 h at room temperature and subsequently developed using Western Lightning Plus-ECL (Perkins Elmer) or SuperSignal West Femto (Thermo).

Immunofluorescence, immunofluorescence–FISH, TIF assay, APB assay, and CO–FISH. Cells grown on coverslips were fixed in 4% paraformaldehyde for 10 min at room temperature. Coverslips were then permeabilized in 0.5% Triton X-100 for 5 min at 4 °C (for most antibodies) or 100% cold methanol for 10 min at –20 °C (for anti-PCNA). Primary antibody incubation was performed at 4 °C in a humidified chamber overnight unless otherwise indicated. Coverslips were washed and incubated with appropriate secondary antibody for 20 min at 37 °C, then mounted onto glass slides using Vectashield mounting medium with DAPI (Vector Labs). For immunofluorescence–FISH, coverslips were re-fixed in 4% paraformaldehyde for 10 min at room temperature after secondary antibody binding. Coverslips were then dehydrated in an ethanol series (70%, 90%, 100%) and allowed to air dry. Dehydrated coverslips were denatured and incubated with TelC–Cy3 peptide nucleic acid (PNA) probe (Panagene) in hybridization buffer (70% deionized formamide, 10 mM Tris (pH 7.4), 0.5% Roche blocking solution) overnight at room temperature in a humidified chamber. The next day, coverslips were washed and mounted as described above. Images were acquired with a QImaging RETIGA-SRV camera connected to a Nikon Eclipse 80i microscope. For TIF assay, cells were scored for co-localized 53BP1 and telomere foci by immunofluorescence–FISH. For APB assay, cells were scored for the number of PML–telomere colocalizations by immunofluorescence–FISH. Hop2 immunofluorescence and CO–FISH experiments were performed as previously described⁷.

Pulsed-field gel electrophoresis and in-gel hybridization. Telomere gels were performed using telomere restriction fragment (TRF) analysis. Genomic DNA was digested using AluI and MboI (NEB). 4–10 μ g of DNA was run on a 1% PFGE agarose gel (Bio-Rad) in 0.5 \times TBE buffer using the CHEF–DR11 system (Bio-Rad) at 6 V cm⁻¹; initial switch time 5 s, final switch time 5 s, for 16 h at 14 °C. The gel was then dried for 4 h at 50 °C, denatured in a 0.5 N NaOH 1.5 M NaCl solution, and neutralized. Gel was hybridized with ³²P-labelled (CCCTAA)₆

oligonucleotides in Church buffer overnight at 42 °C. The next day, the membrane was washed four times in 4 \times SSC buffer, exposed onto a storage phosphor screen (GE Healthcare) and scanned using STORM 860 with ImageQuant (Molecular Dynamics). Telomere length was determined using TeloTool software³⁵.

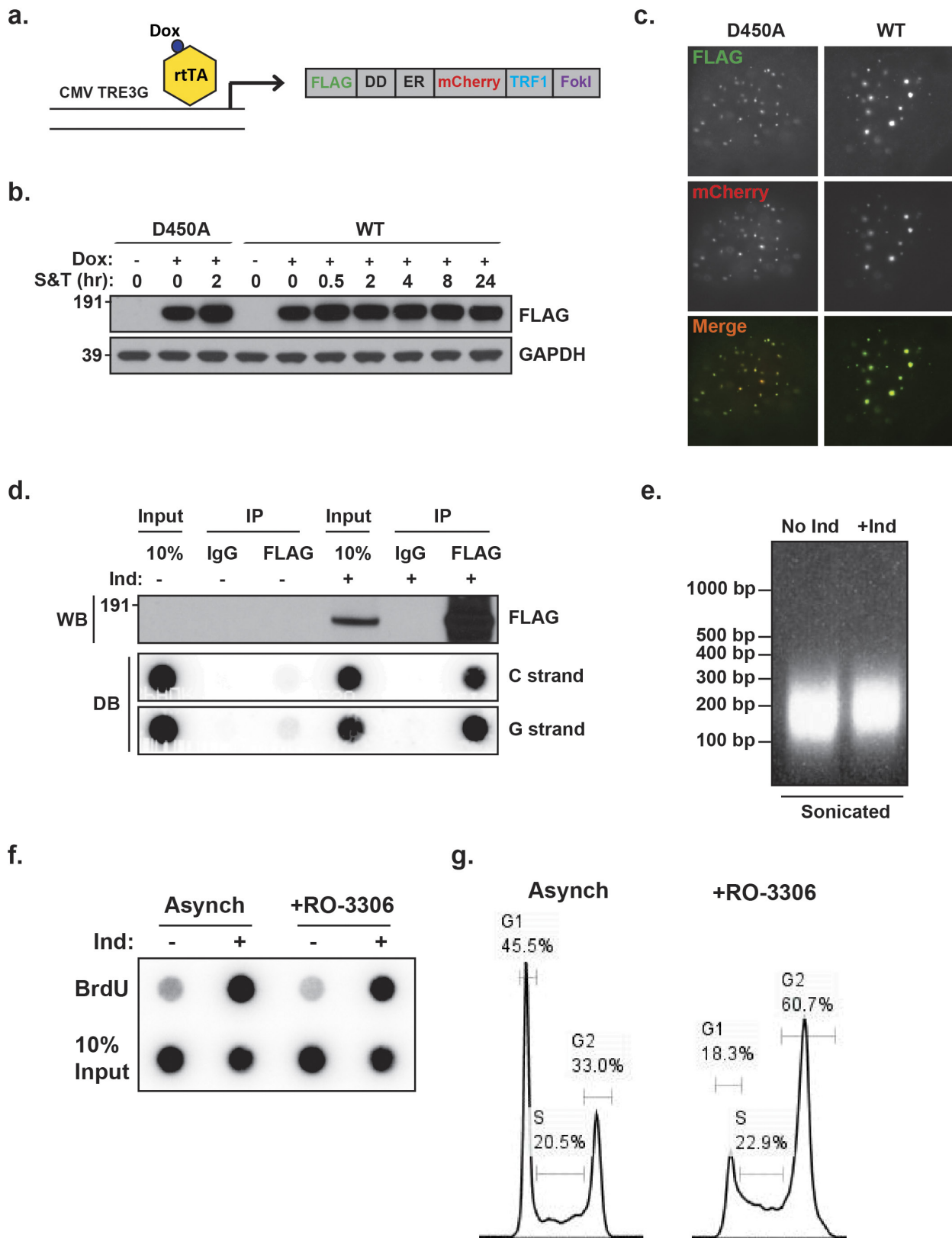
C-circle assay. C-circle assay was performed as previously described²⁸. Genomic DNA was digested using AluI and MboI (NEB). 30 ng of digested DNA was combined with 0.2 mg ml⁻¹ BSA, 0.1% Tween, 1 mM each dNTP without dCTP, 1 \times ϕ 29 Buffer (NEB) and 7.5 U ϕ 29 DNA polymerase (NEB). Samples were incubated for 8 h at 30 °C followed by 20 min at 65 °C. Samples were then diluted in 2 \times SSC buffer and dot-blotted onto an Amersham Hybond-N⁺ nylon membrane (GE). Membrane was ultraviolet crosslinked and then hybridized with ³²P-labelled (CCCTAA)₆ oligonucleotides in PerfectHyb Plus Hybridization Buffer (Sigma) overnight at 37 °C. The next day, the membrane was washed twice in 2 \times SSC buffer, exposed onto a storage phosphor screen (GE Healthcare) and scanned using STORM 860 with ImageQuant (Molecular Dynamics).

Co-immunoprecipitation and chromatin immunoprecipitation (ChIP). Cells were lysed in HEPES immunoprecipitation buffer (10 mM HEPES (pH 8), 2 mM EDTA, 0.1% NP-40) supplemented with 5 mM DTT, 1 mM PMSF, and 1 \times cOmplete protein inhibitor cocktail (Roche) on ice and subsequently spun down at max speed at 4 °C. The supernatant was removed and protein concentration determined using the Protein Assay Dye Reagent (Bio-Rad). 25 μ g protein was removed for input. 500 μ g protein was diluted to 1 mg ml⁻¹ in HEPES immunoprecipitation buffer and pre-cleared with 10 μ l Protein G magnetic beads (Pierce) for 1 h rotating at 4 °C. Protein lysate was then incubated with 10 μ g anti-IgG (Sigma) or anti-POLD1 antibody (Abcam) rotating overnight at 4 °C. The next day, samples were incubated with 30 μ l Protein G magnetic beads (Pierce) that had been pre-bound to a bridging antibody (Active Motif) for 1 h rotating at 4 °C. Beads were subsequently washed five times with HEPES immunoprecipitation buffer. Proteins were eluted by incubating beads with 2 \times sample buffer with BME for 5 min at 95 °C. Samples were analysed by western blot. ChIP was performed as previously described and analysed by western blot and dot blot³⁶.

Telomere content dot blot. 400 ng of genomic DNA was diluted into 2 \times SSC buffer, treated at 95 °C for 5 min, and dot-blotted onto an Amersham Hybond-N⁺ nylon membrane (GE). Membrane was then denatured in a 0.5 N NaOH 1.5 M NaCl solution, neutralized, and UV crosslinked. Membrane was hybridized with ³²P-labelled (CCCTAA)₆, or Alu repeat oligonucleotides in PerfectHyb Plus Hybridization Buffer (Sigma) overnight at 37 °C. The next day, the membrane was washed twice in 2 \times SSC, exposed onto a storage phosphor screen (GE Healthcare) and scanned using STORM 860 with ImageQuant (Molecular Dynamics).

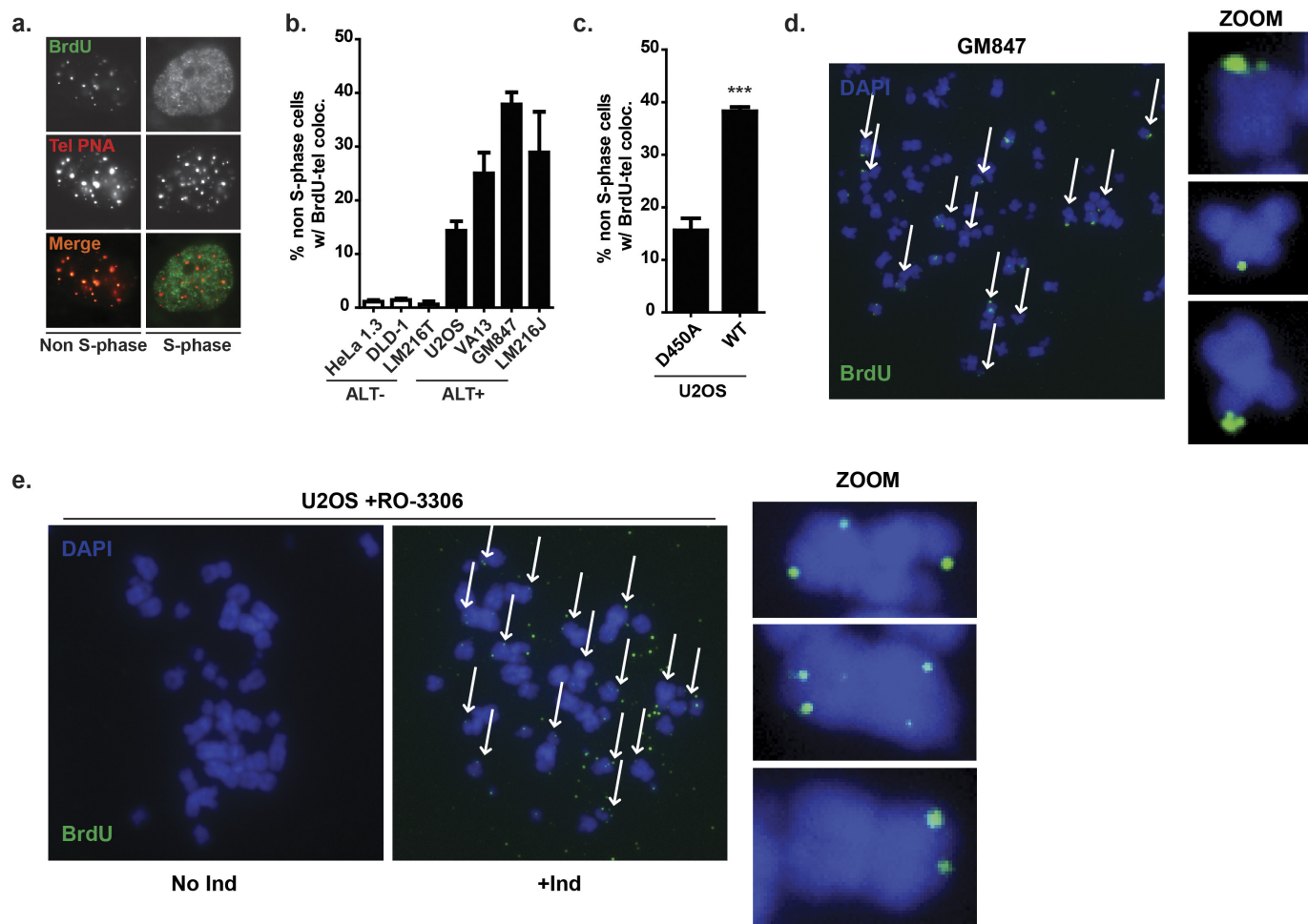
Live cell imaging and image analysis. Live cell imaging was performed and analysed as previously described⁷. Fixed cell and live cell images were captured at 60 \times and 100 \times magnification, respectively. Microscope images and dot blots were prepared and analysed using Fiji. Southern blot telomere gel images were prepared using Fiji and were not cropped to exclude any part of the presented lanes. Western blot gel images were prepared using Adobe Photoshop and cropped to present relevant bands. Uncropped western blot images are shown in Supplementary Fig. 1. **Statistics.** All statistical analysis was done using GraphPad Prism 5 software. Unpaired *t*-tests were used to generate two-tailed *P* values.

34. Viggiani, C. J., Knott, S. R. V. & Aparicio, O. M. Genome-wide analysis of DNA synthesis by BrdU immunoprecipitation on tiling microarrays (BrdU-IP-chip) in *Saccharomyces cerevisiae*. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5385 (2010).
35. Göhring, J., Fulcher, N., Jacak, J. & Riha, K. TeloTool: a new tool for telomere length measurement from terminal restriction fragment analysis with improved probe intensity correction. *Nucleic Acids Res.* **42**, e21 (2014).
36. Tang, J. *et al.* Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. *Nat. Struct. Mol. Biol.* **20**, 317–325 (2013).



Extended Data Figure 1 | An inducible system for studying break-induced telomere synthesis. **a**, Schematic of inducible TRF1-FokI system. **b–d**, Characterization of U2OS inducible TRF1-FokI system by western blot (**b**), immunofluorescence (**c**), and telomere ChIP (**d**). **e**, Agarose gel of sonicated DNA prepared for BrdU pulldown. **f**, **g**, BrdU pulldown dot blot for telomere content (**f**) from asynchronous

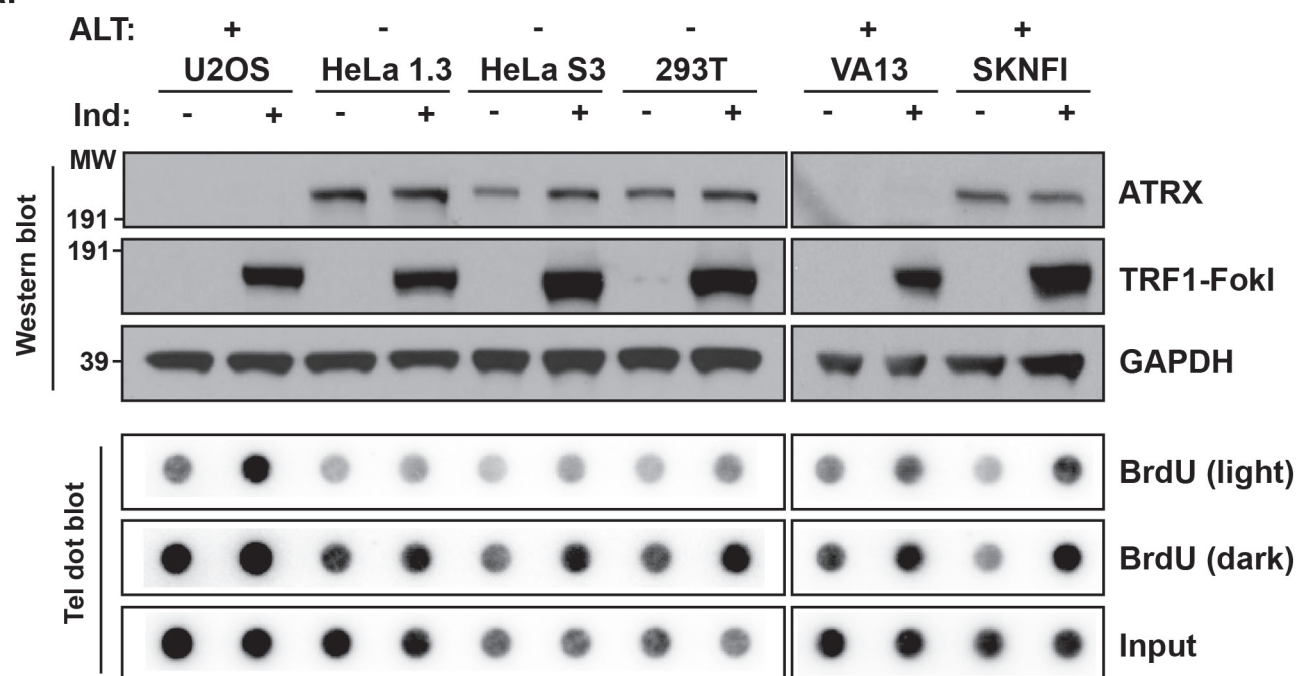
or G2-enriched U2OS cells induced (Ind) with TRF1-FokI for 2 h, with cell-cycle profiles by propidium iodide staining (**g**). Images were captured at 60 \times magnification. Dox, doxycycline; S, Shield-1; T, 4-hydroxytamoxifen; DD, destabilization domain; ER, oestrogen receptor; rtTA, reverse tetracycline transactivator; TRE3G, tetracycline response element; WT, wild-type; D450A, nuclease-null mutant.



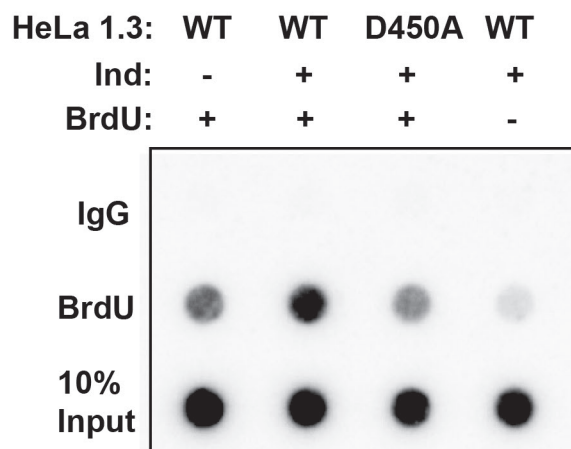
Extended Data Figure 2 | Visualization of spontaneous ALT telomere synthesis. a–c, BrdU immunofluorescence assay to visualize spontaneous ALT telomere synthesis, with representative images of VA13 cells (a) and quantification of a panel of ALT[−] and ALT⁺ cell lines (b) and U2OS cells induced with TRF1–FokI for 2 h (c). d, e, Representative images of BrdU

immunofluorescence of metaphases from spontaneous GM847 cells (d) and U2OS induced (+Ind) with TRF1–FokI for 2 h upon release from RO-3306 (e). Images were captured at 60× magnification. Data represent mean ± s.e.m. of three independent experiments. *** $P \leq 0.001$.

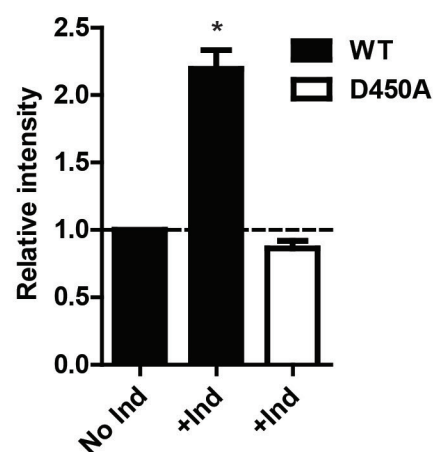
a.



b.

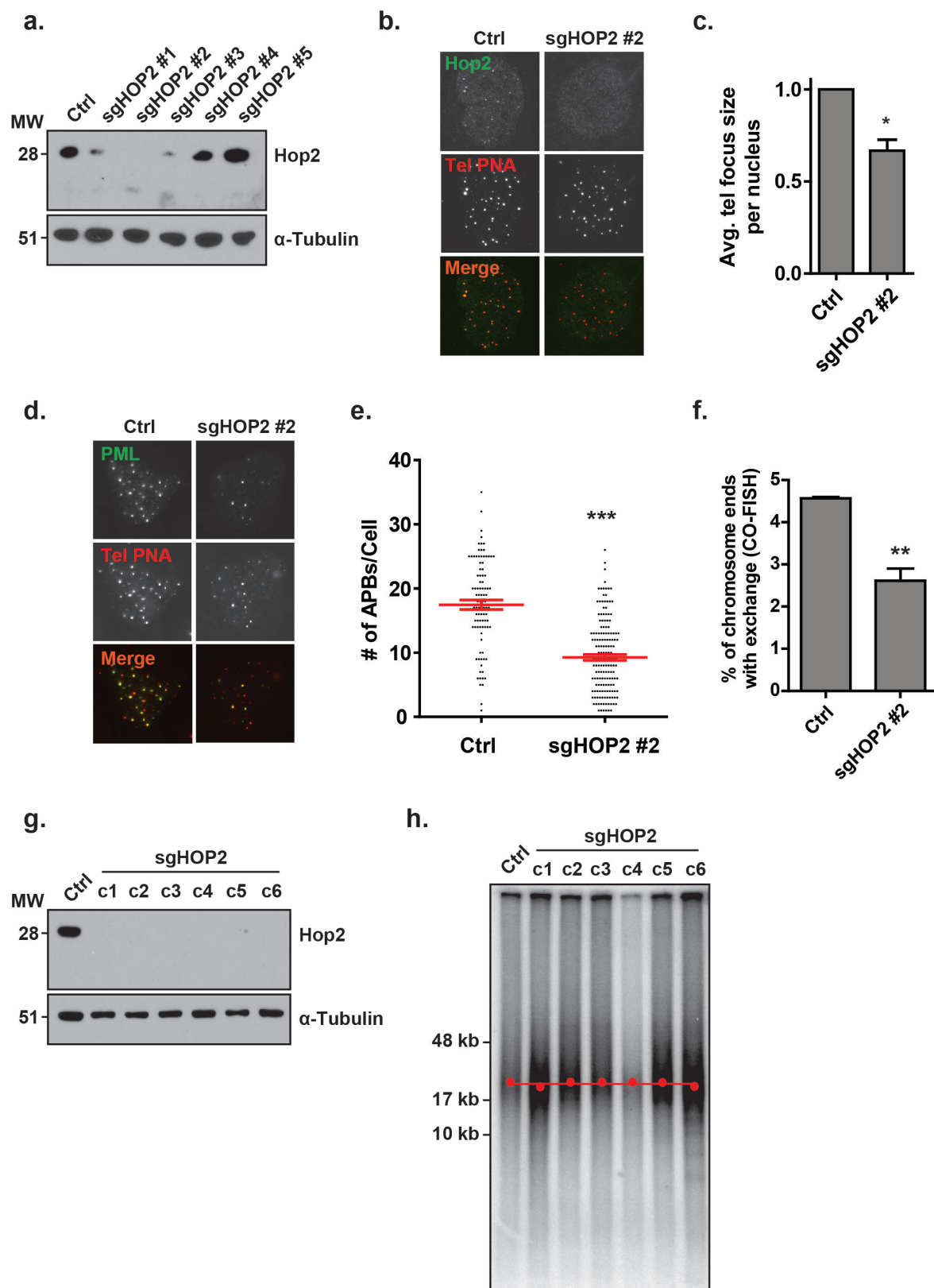


c.



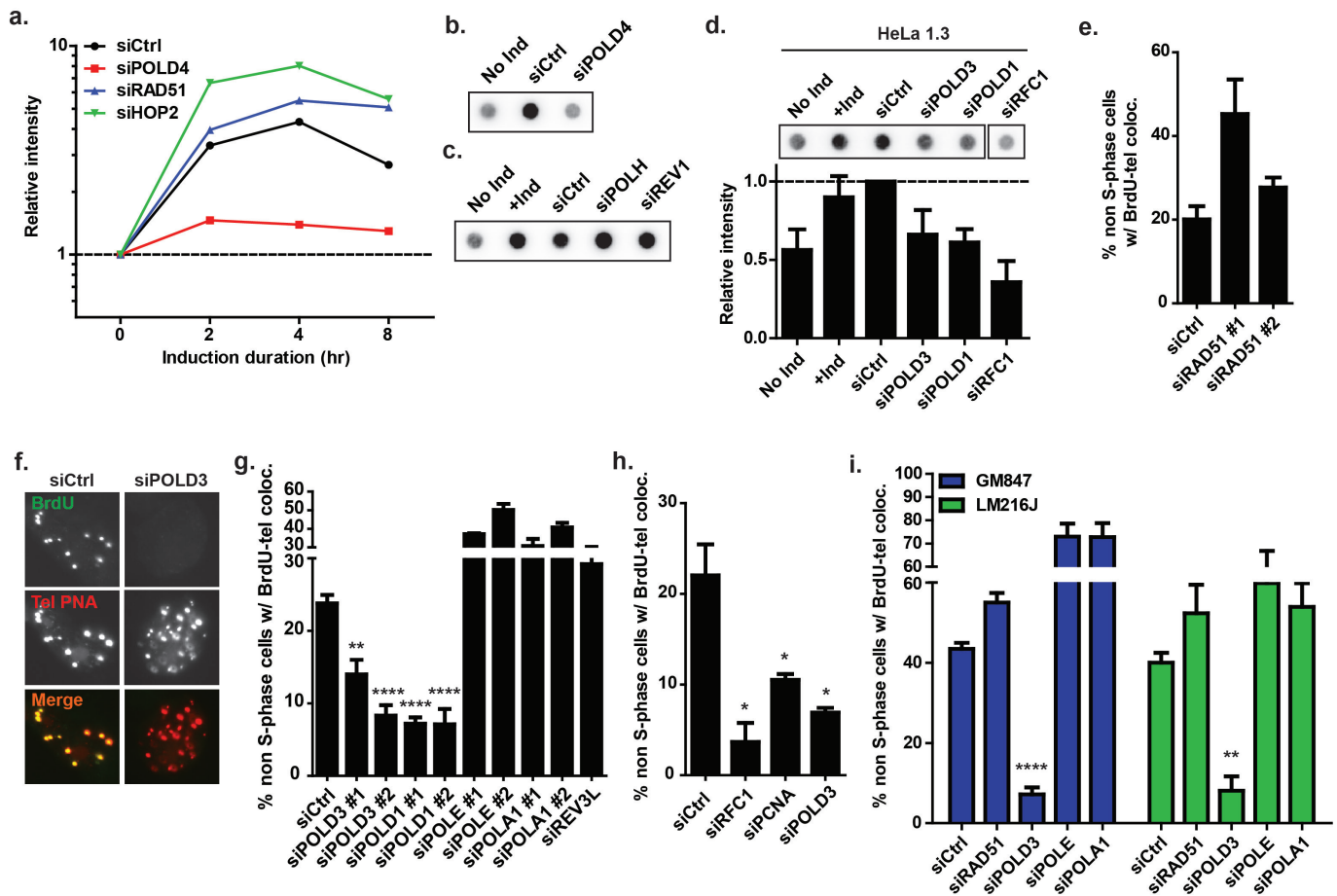
Extended Data Figure 3 | Break-induced telomere synthesis occurs independently of telomere maintenance mechanism. a, A panel of ALT⁻ and ALT⁺ inducible TRF1-FokI cell lines tested for TRF1-FokI and ATRX expression by western blot and nascent telomere synthesis by BrdU pulldown dot blot for telomere content after induction (Ind) with

TRF1-FokI for 2 h. b, c, BrdU pulldown dot blot for telomere content (b) from HeLa 1.3 cells induced (Ind) with TRF1-FokI for 2 h, with quantification (c). Data represent mean \pm s.e.m. of two independent experiments. * $P \leq 0.05$.



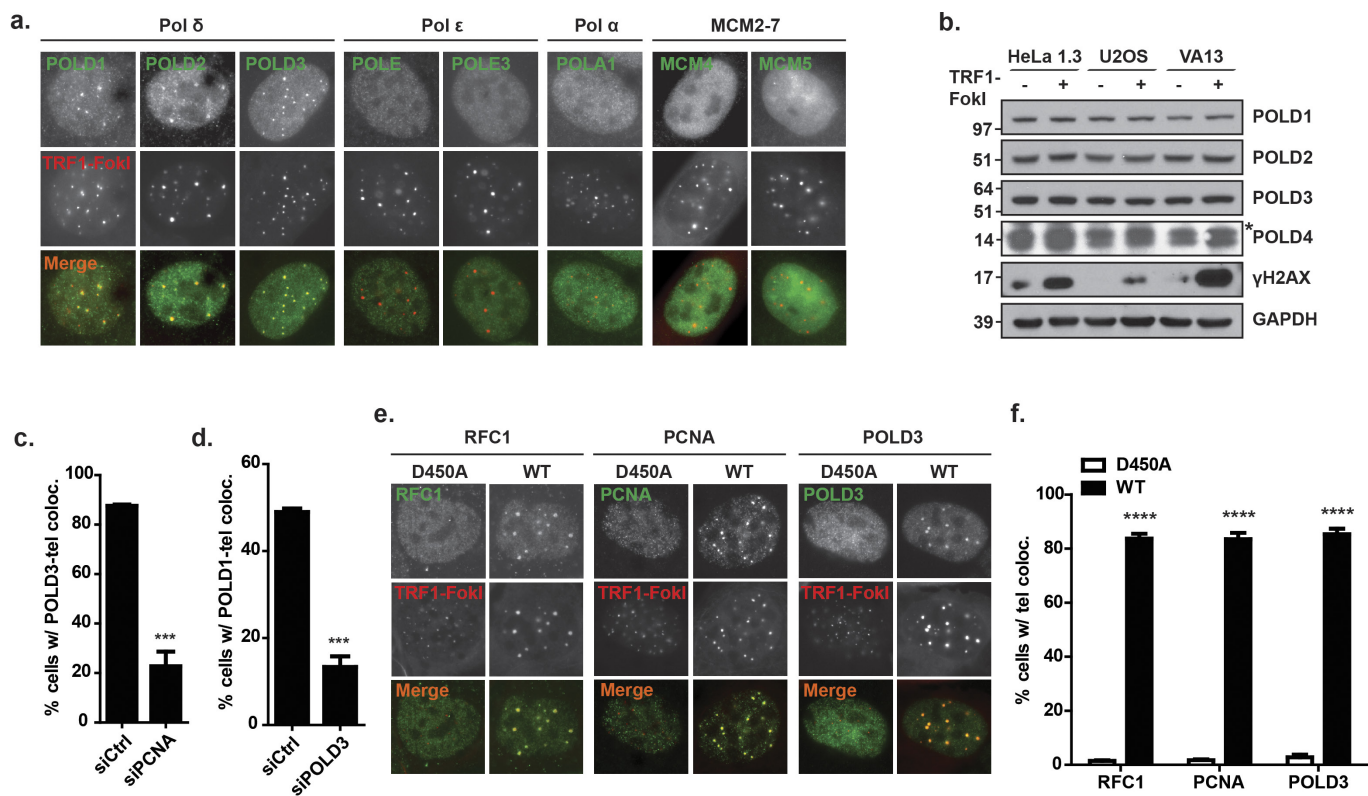
Extended Data Figure 4 | Hop2 contributes to telomere clustering but is dispensable for telomere length maintenance. a–h, CRISPR/Cas9-mediated excision of *HOP2* (sgHOP2) in VA13 cells, with western blot of populations (a). Analysis of Hop2 co-localization with telomere foci by IF-FISH (b), telomere focus size by FISH (c), APBs by PML co-localization with telomere foci (d, e), and telomere exchanges by CO-FISH (f) from

sgHOP2 #2 population. Analysis of clones (c1–c6) by western blot (g) and TRF pulsed-field gel at ~PD 25 (h). Peak intensity of telomere length is indicated by red dot. Images were captured at 60 \times magnification. Data represent mean \pm s.e.m. of at least two independent experiments. *** P < 0.0005, ** P < 0.005, * P < 0.05.



Extended Data Figure 5 | Requirements for break-induced and spontaneous ALT telomere synthesis. **a**, BrdU pull-down dot blot timecourse for telomere content from U2OS induced (Ind) with TRF1-FokI for indicated times and treated with indicated siRNAs. **b**, **c**, BrdU pull-down dot blots for telomere content from U2OS induced (Ind) with TRF1-FokI for 2 h and treated with indicated siRNAs. **d**, BrdU pull-down dot blot for telomere content from HeLa 1.3 induced (Ind) with

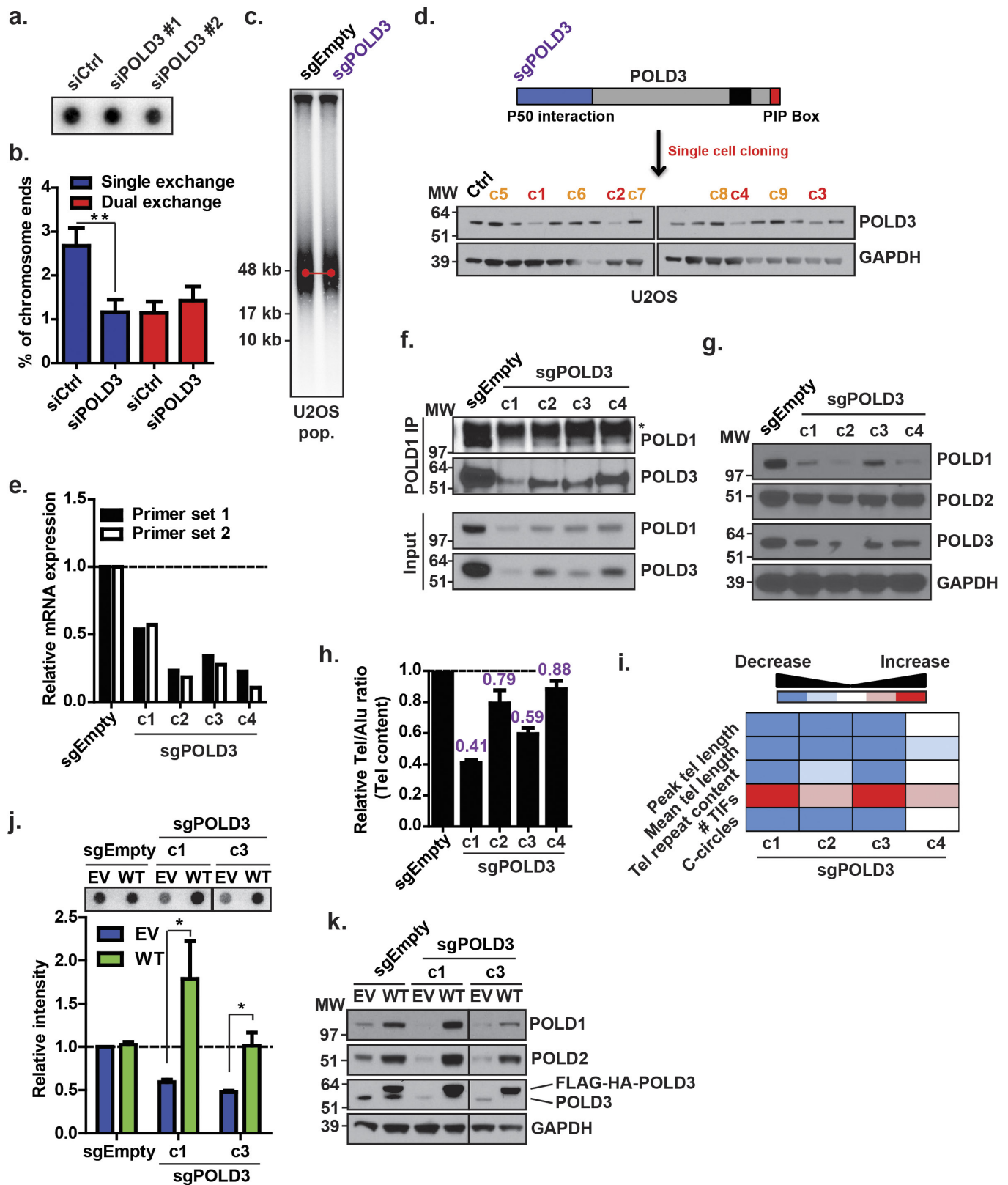
TRF1-FokI for 2 h and treated with indicated siRNAs, with quantification. **e**–**i**, Analysis of spontaneous ALT telomere synthesis using BrdU immunofluorescence from VA13 (**e**–**h**) and GM847 and LM216J (**i**) treated with indicated siRNAs. Images were captured at 60 \times magnification. Data represent mean \pm s.e.m. of two (**d**) or three (**e**–**i**) independent experiments. **** $P \leq 0.0001$, ** $P \leq 0.01$, * $P \leq 0.05$.



Extended Data Figure 6 | Pol δ predominates at ALT telomeres.

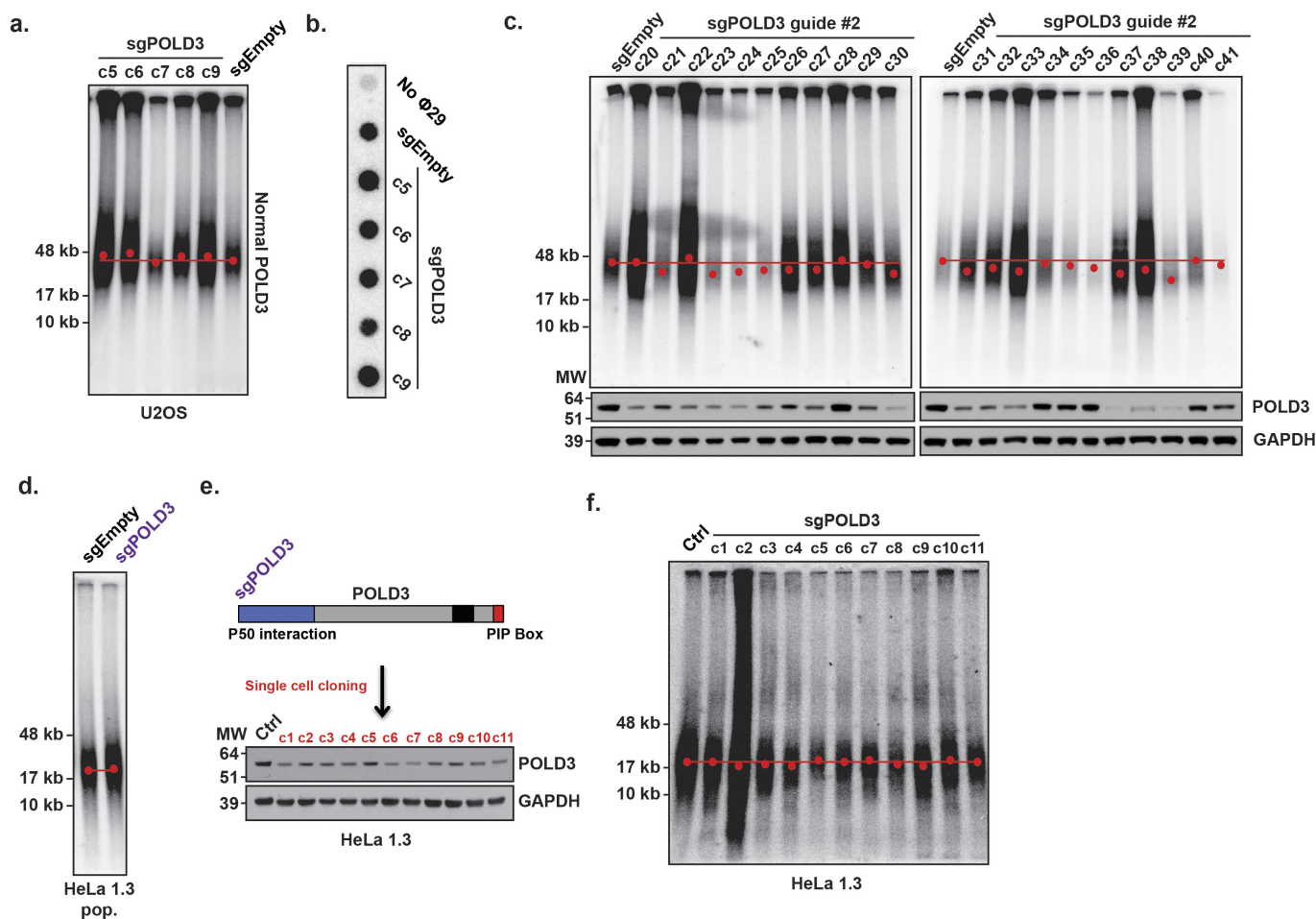
a. Representative images of replisome components (green) and telomere foci (red) from U2OS induced with TRF1-FokI for 2 h. **b.** Western blot of Pol δ complex from cell lines treated with TRF1-FokI. Asterisk denotes non-specific band. **c, d.** Quantification of co-localized POLD3 (**c**) or POLD1 (**d**) with telomere foci from U2OS induced with TRF1-FokI

for 2 h. **e, f.** Representative images (**e**) of co-localized RFC1-PCNA-POLD3 (green) and telomere foci (red) from U2OS induced with TRF1-FokI for 2 h, with quantification (**f**). WT = wild-type, D450A = nuclease-null mutant. Images were captured at 60 \times magnification. Data represent mean \pm s.e.m. of three independent experiments. **** $P \leq 0.0001$, *** $P \leq 0.001$.



Extended Data Figure 7 | POLD3 is critical for telomere maintenance in ALT-dependent cells. **a, b,** analysis of transient POLD3 depletion by C-circle dot blot (**a**) from U2OS and CO-FISH (**b**) from VA13 ($n = 1780$ ends for siCtrl, $n = 1637$ ends for siPOLD3). **c.** TRF analysis from U2OS populations at ~PD 25. Peak intensity of telomere length is indicated by red dot. **d.** Schematic of U2OS POLD3 CRISPR (*sgPOLD3*) cloning strategy with western blot. **e–g.** analysis of POLD3 expression from U2OS clones c1–c4 by qPCR (**e**), POLD1 Co-IP (**f**), and darker exposure of

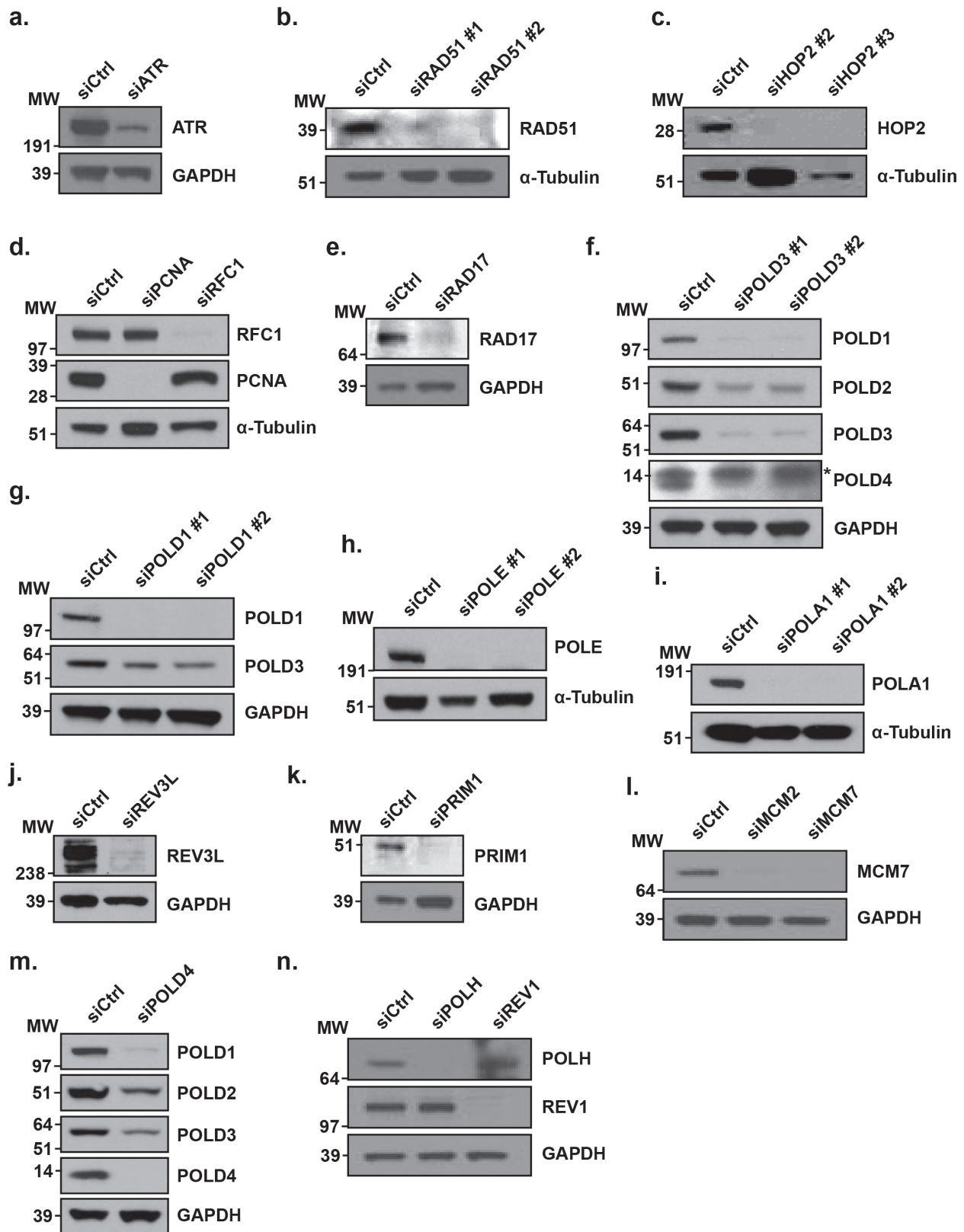
western blot from Fig. 4a (**g**). Asterisk denotes non-specific band. **h.** Quantification of relative telomere content by dot blot from U2OS clones c1–c4. **i.** Heat map summarizing decreases (blue), increases (red), or no change (white) in telomere maintenance from U2OS clones c1–c4 as compared to U2OS control. **j, k.** POLD3-reconstituted CRISPR clones analysed for C-circles by dot blot (**j**) and Pol δ expression by western blot (**k**). EV, empty vector; WT, reconstituted POLD3. Data represent mean \pm s.e.m. of two independent experiments. ** $P \leq 0.01$, * $P \leq 0.05$.



Extended Data Figure 8 | Extended analysis of POLD3 CRISPR clones.

a, b, TRF analysis by pulsed-field gel (**a**) and C-circle dot blot (**b**) from U2OS POLD3 CRISPR (sgPOLD3) clones with normal POLD3 expression (c5–c9) at ~PD 25. **c,** U2OS POLD3 CRISPR clones from an independent guide RNA (sgPOLD3 #2) analysed by TRF and western blot at ~PD 25.

d, TRF analysis by pulsed-field gel from HeLa 1.3 populations at ~PD 25. **e,** Schematic of HeLa 1.3 POLD3 CRISPR cloning strategy with western blot. **f,** TRF analysis by pulsed-field gel from HeLa 1.3 clones c1–c11 at ~PD 25. Peak intensity of telomere length is indicated by red dot.



Extended Data Figure 9 | Knockdown efficiencies. a–n, Western blots of U2OS or VA13 cells treated with indicated siRNAs. Asterisk denotes non-specific band.

Defining synonymous codon compression schemes by genome recoding

Kaihang Wang¹, Julius Fredens^{1*}, Simon F. Brunner^{1*}, Samuel H. Kim^{1,2}, Tionsun Chia¹ & Jason W. Chin^{1,2}

Synthetic recoding of genomes, to remove targeted sense codons, may facilitate the encoded cellular synthesis of unnatural polymers by orthogonal translation systems. However, our limited understanding of allowed synonymous codon substitutions, and the absence of methods that enable the stepwise replacement of the *Escherichia coli* genome with long synthetic DNA and provide feedback on allowed and disallowed design features in synthetic genomes, have restricted progress towards this goal. Here we endow *E. coli* with a system for efficient, programmable replacement of genomic DNA with long (>100-kb) synthetic DNA, through the *in vivo* excision of double-stranded DNA from an episomal replicon by CRISPR/Cas9, coupled to lambda-red-mediated recombination and simultaneous positive and negative selection. We iterate the approach, providing a basis for stepwise whole-genome replacement. We attempt systematic recoding in an essential operon using eight synonymous recoding schemes. Each scheme systematically replaces target codons with defined synonyms and is compatible with codon reassignment. Our results define allowed and disallowed synonymous recoding schemes, and enable the identification and repair of recoding at idiosyncratic positions in the genome.

The design and synthesis of genomes provides a powerful approach for understanding and engineering biology^{1–6}. Genome synthesis has the potential to elucidate synonymous codon function⁷, accelerate metabolic engineering⁸, and facilitate genetically encoded unnatural polymer synthesis^{9,10}.

Methods that replace the genome in sections⁶, provide feedback on precisely where a given design fails and on how to repair it, and can be rapidly iterated for whole-genome replacement, would increase our ability to understand and manipulate the information encoded in genomes.

In *E. coli*, the workhorse of synthetic biology, progress on replacing large sections of the genome has been slower than in naturally recombinogenic organisms^{6,11}. Sequence-specific recombinases may be introduced into *E. coli* to direct recombination at defined target sequences that must be introduced into the genome in advance, and these approaches cannot be iterated¹². Lambda-red-mediated homologous recombination¹³, using linear double-stranded (ds)DNA that is electroporated in to cells, can be programmed to target any region of the genome via short (50-bp) homology regions at either end of a linear dsDNA (referred to herein as HR1 and HR2). However, this approach is commonly limited to inserting or replacing only 2–3 kb of genomic DNA, and has not been used to introduce long sequences into the genome.

We are interested in reprogramming the genetic code for the *in vivo* biosynthesis of unnatural polymers⁹. Reassigning particular codons in the genome to synonymous codons would enable removal of their cognate transfer RNAs, compression of the number of synonymous codons used to encode certain natural amino acids, and the reassignment of certain sense codons, and an expanded set of quadruplet codons to evolved orthogonal translation systems for unnatural polymer synthesis^{14,15}. However, recoding the *E. coli* genome requires both the development of methods for efficiently replacing genomic DNA with synthetic DNA and an understanding of the best synonymous codon substitutions, from many possible choices, for recoding.

Nature chooses one triplet codon from up to six potential synonyms to encode each amino acid at each position in the genome; this choice

can define transcriptional¹⁶ or translational¹⁷ regulatory elements, translation speed^{18,19}, mRNA folding⁷, gene expression, co-translational folding^{20,21} and protein production levels⁷, and is likely to have further undiscovered roles. Synonymous codons may have distinct roles at different sites in the genome, and there may be epistatic interactions amongst codons within and between genes^{22–24}. Our limited understanding of the factors driving codon choice suggests that the best synonymous codon substitutions to implement for synthetic recoding should be identified empirically.

Here we endow *E. coli* with a system that enables efficient, programmable, one-step introduction of long synthetic DNA into the genome, as insertions or replacements, and iterate the approach for stepwise replacement of longer genomic regions. Using our approach we investigate different synonymous recoding schemes for replacing the same target codons with distinct sets of synonyms, in an operon rich in both target codons and essential genes, providing insight into allowed and disallowed schemes for genome recoding and synonymous codon compression.

Inserting DNA into the genome by REXER

The overall efficiency of lambda-red-mediated recombination protocols is the product of the transformation efficiency for linear dsDNA and the efficiency with which the linear dsDNA mediates intracellular recombination. The overall efficiency decreases markedly with increasing dsDNA length, and we hypothesized that this decrease results primarily from challenges in efficiently delivering intact dsDNA into cells. To address this challenge, we envisioned introducing the DNA of interest into *E. coli* in an episomal replicon, and excising the dsDNA of interest to facilitate lambda-red-mediated recombination. To select for the correct integrants we envisioned the use of simultaneous positive and negative selection, to select for the integration of a positive selection marker from the replicon into the genome and the loss of a negative selection marker from the genomic locus targeted for replacement; such a double selection strategy substantially enhances integration at the target locus by lambda-red-mediated recombination (Extended Data Fig. 1).

¹Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK. ²Department of Chemistry, Cambridge University, Cambridge CB2 1EW, UK.

*These authors contributed equally to this work.

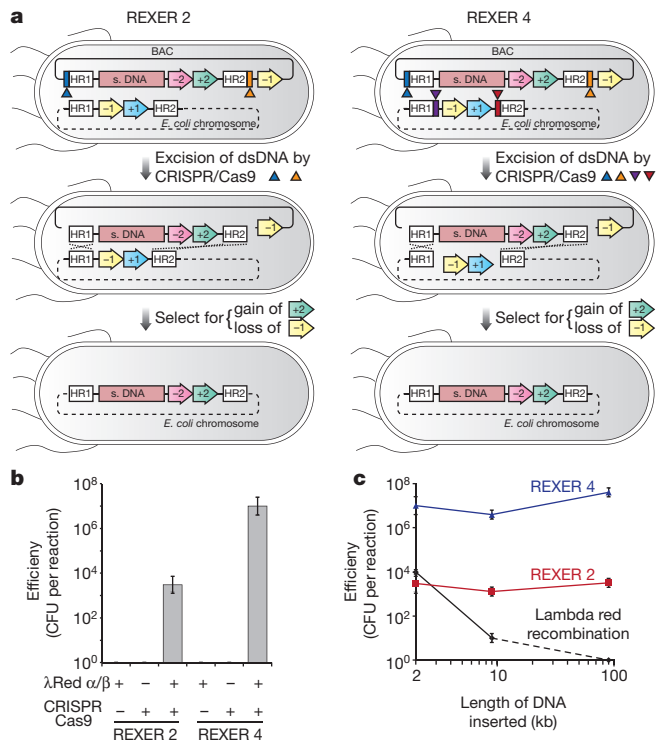


Figure 1 | Efficient, programmable insertion of very long synthetic DNA (s. DNA) into the genome of *E. coli*. **a**, REXER 2 and REXER 4. CRISPR protospacer sequences are blue and orange rectangles, respectively. Triangles indicate spacer RNAs that program cleavage within colour-matched protospacers. REXER 4 augments REXER 2 by adding two extra protospacers (purple and red rectangles) and triggering cleavage with four spacer RNAs. +1 is *Kan^R*, -1 is *rpsL*, +2 is *Cm^R*, -2 is *sacB*. **b**, REXER 2 and REXER 4 are dependent on the CRISPR/Cas9 system and recombination. For experiments that omit CRISPR/Cas9 or the recombination machinery, the efficiency is significantly reduced ($P < 10^{-4}$, two tailed *t*-test, for the null hypothesis that the efficiency is independent of CRISPR/Cas9 or lambda red for REXER 2 and REXER 4). Controls omit either spacer RNA or lambda red beta. Data show mean \pm s.d. ($n = 4$, two biological replicates with two technical replicates of each for experiments that omit lambda red alpha/beta (λ Red α/β) or CRISPR/Cas9; $n = 6$, three biological replicates with two technical replicates of each for other experiments). **c**, The efficiency of REXER 2 and REXER 4 is constant for insertions between 2 kb and 90 kb. CFU, colony forming units. The data show the mean \pm s.d. ($n = 6$, three biological replicates with two technical replicates of each for 2-kb insertion; $n = 4$, two biological replicates with two technical replicates of each for 9-kb and 90-kb insertions). It was not possible to obtain a 90-kb linear dsDNA product *in vitro* for classical lambda red recombination, and our data reflect this, rather than the efficiency of recombination *per se*. It is well established that lambda red recombination efficiency falls off rapidly with linear dsDNA length.

We created *E. coli* MDS42^{rpsLK43R/rK} (a genome-minimized strain of *E. coli*²⁵ in which the genomic copy of *rpsL* contains a K43R mutation conferring resistance to streptomycin, and the -1/+1 selection cassette encoding an *rpsL*-*Kan^R* fusion inserted between *cra* and *mraZ*) containing a bacterial artificial chromosome (BAC) in which the -2/+2 cassette (encoding *sacB*-*Cm^R*) is flanked by the HR1 and HR2 sequences and Cas9 target sites (containing protospacer-PAM (protospacer adjacent motif) sequences) and expressing lambda red (alpha/beta/gamma)¹³, Cas9²⁶, and *tracrRNA*²⁶ (Fig. 1a). The addition of a plasmid encoding spacer RNAs targeting the protospacers²⁶ within the BAC target sites to these cells, and selection for the gain of resistance to both chloramphenicol (gain of +2) and streptomycin (loss of -1 from the genome, and loss of the backbone of the BAC) led to replacement of the sequence between HR1 and HR2 in the genome,

with the sequence between HR1 and HR2 from the BAC (Fig. 1a and Extended Data Fig. 2).

Genomic replacement was strictly dependent on CRISPR/Cas9 and the lambda red recombination machinery (Fig. 1b), and targeted to the desired genomic locus (Extended Data Fig. 1c, d); consistent with the CRISPR/Cas9-mediated excision of the dsDNA between HR1 and HR2 in the BAC and lambda-red-mediated integration of this sequence between HR1 and HR2 in the genome. We named our approach REXER 2 (replicon excision for enhanced genome engineering through programmed recombination; '2' indicates the number of CRISPR/Cas9 cuts).

To investigate the dependence of REXER 2 on the length of DNA inserted into the genome, we created BACs with 9 kb or 90 kb of DNA inserted between HR1 and -2/+2 (Fig. 1c and Extended Data Fig. 2). The insertions contain a *luxABCDE* operon²⁷, which is sufficient to generate bioluminescence in *E. coli*. We transformed each BAC into *E. coli* MDS42^{rpsLK43R/rK} and implemented the REXER 2 protocol. All cells selected on chloramphenicol and streptomycin integrated the *lux* operon at the correct locus and were bioluminescent (Extended Data Fig. 2). Moreover, whereas the efficiency of classical recombination drops markedly from 10^4 colony forming units (CFU) for a 2-kb insertion to approximately 10 CFU for a 9-kb insertion, the efficiency of REXER 2 is constant, at 10^4 CFU for all insertions (Fig. 1c).

Next we investigated whether the efficiency of REXER could be improved by making two additional CRISPR/Cas9-mediated double-strand breaks in the genome between HR1 and HR2 (Fig. 1a). The resulting REXER 4 protocol led to replacement of the sequence between HR1 and HR2 in the genome with the sequence between HR1 and HR2 from the BAC (Extended Data Fig. 2) and destruction of all four Cas9 target sites. REXER 4 was strictly dependent on both the CRISPR/Cas9 system and the lambda red recombination machinery (Fig. 1b), and led to integration at the correct locus (Extended Data Fig. 2). Like REXER 2, the efficiency of REXER 4 was independent of length for insertions tested (up to 90 kb), and REXER 4 further increased the efficiency of synthetic DNA insertion with respect to REXER 2 by 10^3 -fold while maintaining insertion at the correct locus (Fig. 1c and Extended Data Fig. 1c, d).

Replacing genome sections by REXER

Next, we demonstrated that REXER could be used to efficiently replace 100 kb of the *E. coli* genome in a single step. We targeted the region from *mraZ* to *pyrH* for replacement and inserted the -1/+1 selection cassette at the 5' end of this region. We assembled a BAC from DNA fragments in *Saccharomyces cerevisiae*²⁸ in which the 100-kb region between *mraZ* and *pyrH* was watermarked along its length by genes from the *lux* operon (Extended Data Fig. 3). REXER 2 yielded 2×10^4 CFU per reaction, of which 80% were bioluminescent, and REXER 4 yielded 5×10^6 CFU per reaction, of which 50% were bioluminescent (Extended Data Fig. 3). Further characterization confirmed the integration of the *lux* watermarks at the expected loci for all bioluminescent colonies (Extended Data Fig. 3). These results demonstrate that REXER enables the replacement of genomic regions with synthetic DNA.

Iterative REXER for genome replacement

Iteratively replacing large sections of the genome with synthetic DNA via REXER will enable genome stepwise interchange synthesis (GENESIS) for whole-genome replacement (Fig. 2a). Towards this goal, we demonstrated iterative REXER (Extended Data Fig. 4). The genome created in a first round of REXER, which introduces -2/+2, provides a direct template for a second round of REXER, using a BAC that contains distinct positive and negative selection markers (-1/+1) (Extended Data Fig. 4). This product is a template for a third round of REXER; thus, the approach can be iterated (Extended Data Fig. 4).

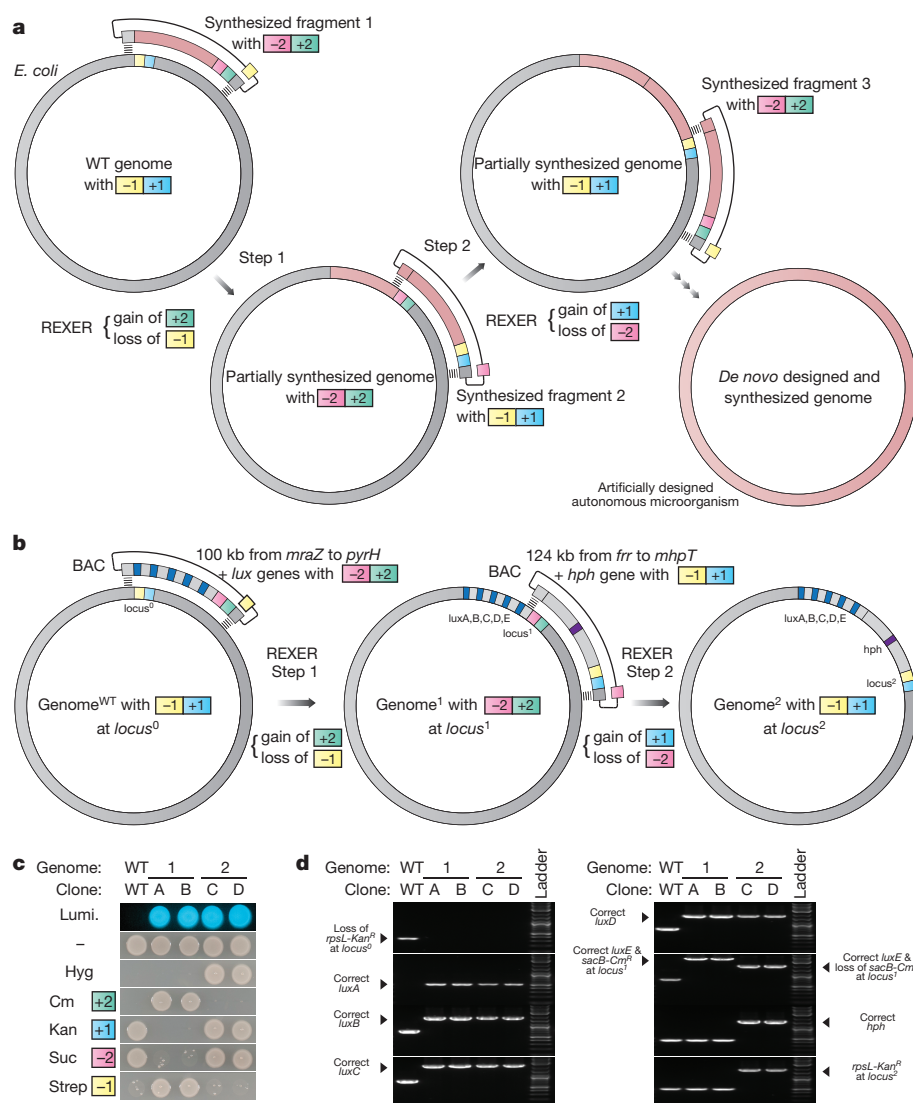


Figure 2 | Iterating REXER for GENESIS. **a**, Iterative genomic replacement by REXER will enable genome replacement in a series of steps. **b**, Iterative REXER replaces 220 kb of the *E. coli* genome with 230 kb of synthetic DNA in two steps. *LuxA*, *B*, *C*, *D* and *E* (cyan rectangles) are necessary and sufficient for luminescence. *hph* (violet rectangle) is the hygromycin B phosphotransferase gene, which confers resistance to hygromycin B. **c**, Cells phenotype correctly through rounds of REXER. genome^{WT}, parental cell line; clones A and B, independent clones from the first round of REXER; clones C and D, independent clones from the second round of REXER; Lumi, luminescence; Cm, chloramphenicol; Kan, kanamycin; Suc, sucrose; Strep, streptomycin. **d**, Cells genotype correctly through rounds of REXER. For gel source images, see Supplementary Fig. 1.

To explicitly demonstrate the stepwise replacement of long regions of genomic DNA with synthetic DNA for GENESIS, we used cells in which we replaced the 100-kb genomic region between *mraZ* and *pyrH* by REXER (Extended Data Fig. 3 and Fig. 2b) for a second step of REXER. This second step introduced 124 kb of DNA spanning *frr* to *mhpT* and the hygromycin B phosphotransferase gene (*hph*). We confirmed the replacement of 220 kb of the genome with 230 kb of synthetic DNA in two steps (Fig. 2c, d). This compares favourably with the largest replacement in the naturally recombinogenic *S. cerevisiae* (270 kb, 11 steps)⁶.

Testing synonymous recoding schemes

We used REXER to define synonymous substitutions that are disallowed and poorly tolerated and synonymous substitutions that are allowed and can be implemented at many positions in the genome. To define a system for experimental investigation we identified: i) the codons to target for removal; ii) the codons with which the target codons might be replaced to define recoding rules; and iii) a region of the genome in which to test recoding rules.

We chose target codons that, when removed from the genome, would enable the removal of all the tRNAs that decode them, and where removal of these tRNAs would not remove all decoding of the remaining synonymous codons in the genome; these are the minimum criteria for removing a sense codon from the genome to enable its unambiguous reassignment (Extended Data Fig. 5). We focused on removing

serine, leucine and alanine codons that fulfill these criteria, as these are the three codon sets for which the aminoacyl-tRNA synthetases do not recognize the anticodons of their cognate tRNAs²⁹. This will facilitate reassignment of the target codon (or up to four quadruplet derivatives), following deletion of host tRNAs, to orthogonal synthetase-tRNA pairs in orthogonal translation¹⁴.

We defined candidate synonymous replacements for the target codons by identifying the closest match for the target codons, as judged by either codon adaptation index (cAi)³⁰, tRNA adaptation index (tAi)^{31,32}, or a third metric we define (translation efficiency, tE) (Extended Data Table 1 and Methods). These considerations led us to investigate eight recoding schemes (Fig. 3a).

We identified the *E. coli* cell division operon as an ideal target in which to test these synonymous recoding schemes because it i) is rich in essential genes (12 out of 15 genes in the region are essential)³³, ii) contains genes expressed at a range of levels³⁴, iii) includes genes encoding membrane proteins³⁴ (a class of proteins for which co-translational folding is known to be affected by synonymous codon choice), iv) includes several proteins that interact and for which the ratios of proteins expressed are distinct and crucial^{35,36} (which will favour inter-genetic epistatic interactions amongst codons), and v) is rich in the target codons (Fig. 3b and Extended Data Table 2). We anticipated that these features would ensure that the region captures important properties of the genome, and that the success or failure of synonymous recoding in the region would be reflected in viability and growth or a lack thereof.

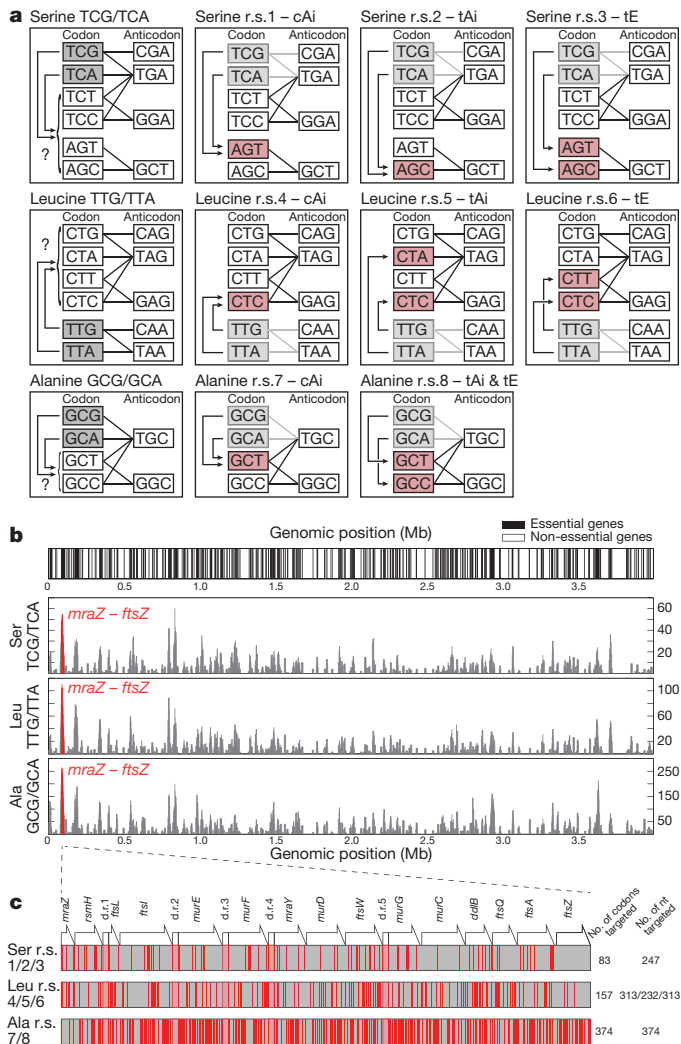


Figure 3 | Systematic and defined synonymous recoding in an *E. coli* operon rich in essential genes. **a**, Identifying codon targets for removal (grey) and the synonyms to which they are recoded (pink) in each recoding scheme (r.s.). Lines indicate codon–anticodon interactions. Replacements were chosen by cAi, tAi, or tE. Application of each recoding scheme genome-wide would allow the targeted codons to be completely removed from the *E. coli* genome and, following deletion of the cognate tRNA genes, codon reassignment to orthogonal translation systems for unnatural polymer synthesis. **b**, Identifying a target operon rich in target codons and essential genes to test recoding schemes. The top panel indicates the positions of essential genes. In the bottom three panels, the *y* axis scores the number of the indicated target codons in essential genes at the genomic position indicated on the *x* axis. The *mraZ* to *ftsZ* region (red) was identified in the highest-scoring 20-kb region across the *E. coli* MDS42 genome for all targeted codons. **c**, Position and density of targeted codons in the *mraZ* to *ftsZ* region. The positions of targeted codons (the indicated sense codons plus TAG to TAA) are coloured in red. Pink regions with red outlines indicate duplicated regions, which refactor² overlapping open reading frames to enable independent recoding of the downstream open reading frames.

Allowed and disallowed recoding schemes

We designed DNA sequences in which each of the recoding schemes was implemented within all of the fifteen genes simultaneously. Overall, the schemes investigate the consequences of 1,468 codon changes and 2,347 nucleotide changes (Fig. 3c). The DNA for each scheme was synthesized *de novo* and assembled into a BAC in *S. cerevisiae*, and genomic recoding via REXER was investigated (Extended Data Figs 6, 7).

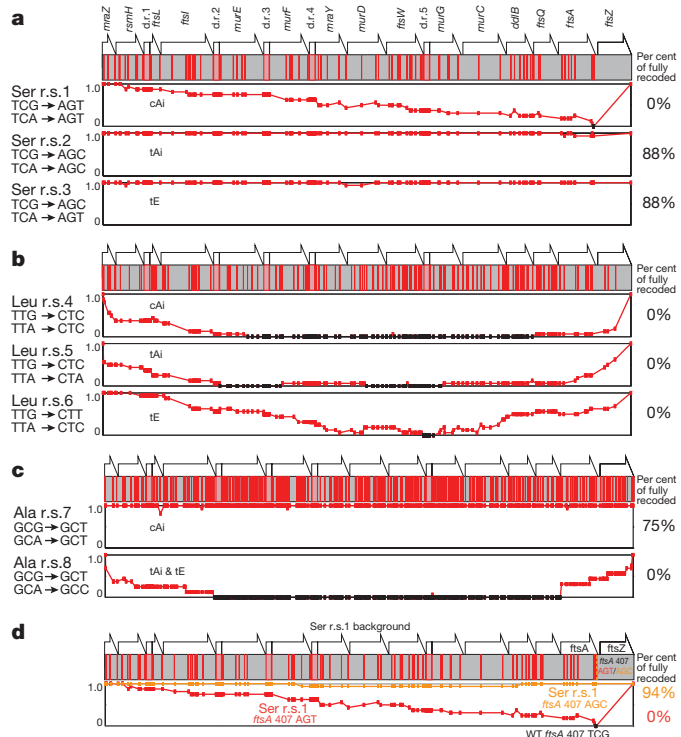


Figure 4 | Compiled recoding landscapes of targeted codons reveal allowed and disallowed synonymous recoding schemes and enable the identification and repair of idiosyncratic positions in the genome. The fraction of recoding across 16 independent sequences is indicated on the y axis of the graphs. Codon positions that are not recoded with the indicated scheme are in black. **a–c**, Compiled recoding landscapes of targeted serine, leucine and alanine codons respectively. **d**, Identifying and fixing a deleterious sequence in defined and systematic synonymous recoding. The compiled recoding landscape of serine recoding scheme 1 is plotted in red, revealing the single position at which the wild-type sequence is maintained (codon 407 in *ftsA*). The compiled recoding landscape of serine recoding scheme 1 with *ftsA* 407 AGT changed to AGC (as in serine recoding schemes 2 and 3) is plotted in orange. This mutation repairs the deleterious effect of *ftsA* 407 AGT without reintroducing the codons targeted for removal.

Following REXER, we sequenced 16 independent clones from each recoding scheme. We observed chimaeras between the wild-type genomic DNA and the recoded DNA in several cases, consistent with recombination-mediated crossover between the recoded sequence and the starting genome; these chimaeras defined a recoding landscape. We aligned the individual recoding landscapes to create a ‘compiled recoding landscape’ (Extended Data Figs 6, 7) that reveals peaks and plateaus for synonymous substitutions that are allowed and valleys or troughs for synonymous substitutions that are consistently disallowed. We observe clear differences in the extent to which replacement of the same target codons by different synonymous codons are tolerated (Fig. 4a–c and Extended Data Figs 6, 7).

We first investigated the serine recoding schemes 1–3 (Fig. 4a). For scheme 1, 0% of clones were completely recoded, but for scheme 2 and scheme 3, 88% of clones were fully recoded. By contrast, none of the leucine recoding schemes tested (schemes 4–6) led to complete recoding, and for schemes 4 and 5 recoding failed catastrophically, indicating that the synonymous substitutions had phenotypic consequences at many sites in the operon (Fig. 4b). Finally, the two alanine recoding schemes tested (schemes 7 and 8) had markedly different outcomes (Fig. 4c). Recoding scheme 7 led to 75% of clones being completely recoded at all 374 positions, whereas no clones were fully recoded by scheme 8. The doubling times for all fully recoded clones were comparable to each other and to a control *E. coli* strain (Extended Data Fig. 8).

Overall, this work successfully removed up to 373 sense codons across 20 kb from an operon rich in essential genes in a single strain. Thus, the scale of sense codon removal was much greater than in previously reported work that investigated one gene at a time³⁷.

Our data reveal the marked differences between precisely defined recoding schemes that are obscured when the choice of synonymous substitution is varied³⁷. For serine recoding, recoding by schemes 2 and 3 is allowed, while scheme 1 recoding is not, even though the codons used for replacement in scheme 1 differ from those used in schemes 2 and 3 by only a single base (AGT vs AGC), and are decoded by the same tRNA (with anticodon GCT) via wobble and Watson–Crick decoding, respectively (Fig. 3a). Similarly, for alanine codons, scheme 7 recoding is allowed while scheme 8 recoding fails catastrophically. These recoding schemes differ only in the conversion of a single base (GCT vs GCC) in the allowed and disallowed substitution for GCA. Again, both of the new codons are decoded by the same set of tRNAs (Fig. 3a). cAI, tAi and tE all produce at least one successful recoding, but no single metric predicts which synonymous recoding will be successful. These observations underscore the importance of empirically determining the best systematic and well-defined synonymous recoding scheme for each codon.

E. coli consistently rejects a single-codon mutation (TCG to AGT) at codon 407 of *ftsA* in recoding scheme 1 (Fig. 4a). Attempts to introduce the *ftsA* 407 TCG-to-AGT mutation (without additional recoding at other positions in the genome) failed (Extended Data Fig. 8). By contrast, we were able to quantitatively recode *ftsA* 407 TCG to the synonymous TCT codon (Extended Data Fig. 8). These results demonstrate that the *ftsA* 407 TCG-to-AGT mutation is deleterious.

Mutation of the codon at position 407 in *ftsA* from AGT to AGC (the codon found at this position in recoding schemes 2 and 3) is sufficient to repair recoding scheme 1 (Fig. 4d and Extended Data Figs 7, 8). This mutation markedly alters REXER-mediated recoding, increasing the fraction of fully recoded sequences from 0% to 94% and the fraction of recoding at codon 407 of *ftsA* from 0% to 100% (Extended Data Fig. 8). We also successfully introduced this mutation into recoding scheme 1 by combining single-stranded DNA recombineering with REXER (Extended Data Fig. 8). The growth of *E. coli* was not affected by the successful recoding schemes (Extended Data Fig. 8). These results demonstrate that the major defect in recoding scheme 1 results from AGT being disallowed at position 407 of *ftsA*, though it is tolerated at many other positions. Since TCG, TCT and AGC are allowed at position 407 of *ftsA* but AGT (which shares nucleotides with allowed codons at each position of the triplet) is disallowed, we conclude that the problem at this codon lies in the entire triplet. These experiments exemplify how REXER may be used to identify idiosyncratic positions in the genome that are refractory to recoding by otherwise well-tolerated recoding schemes, and repair the recoding scheme by the introduction of alternative codons at these idiosyncratic positions.

Discussion

We have generated an efficient approach to enable both the programmed insertion of large synthetic DNA sequences into the *E. coli* genome and the replacement of sections of the *E. coli* genome with synthetic DNA. The approach can be iterated and may enable replacement of the entire *E. coli* genome with synthetic DNA in approximately fourteen steps (Fig. 2a), given the length independence of REXER, demonstrated herein, and the ability of *E. coli* to accept 300-kb BACs. Each step takes only a few days to implement and convergent syntheses may further accelerate complete genome synthesis. The strategy will enable radical, high-density changes to the genome that are not accessible through site-directed mutagenesis approaches^{10,38}, and enable diverse applications including recoding and metabolic engineering. We anticipate that this approach may be extended to facilitate genome engineering in other organisms.

We have simultaneously recoded several genes in an essential operon using eight well-defined synonymous recoding rules that are

compatible with codon reassignment for unnatural polymer synthesis. Our results reveal marked differences in the extent to which different synonymous replacements for the same target codons are allowed. Our approach also enables both the identification and repair of idiosyncratic positions within the ‘recoding landscape’ where a precise codon substitution that is allowed at many other positions in the operon is disallowed. Our investigation empirically identifies precisely defined schemes for sense codon removal and synonymous replacement for genome-wide application.

Note added in proof: Additional work on variable codon substitution³⁷ was recently reported^{39,40}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 February; accepted 26 September 2016.

Published online 24 October 2016.

1. Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* **297**, 1016–1018 (2002).
2. Chan, L. Y., Kosuri, S. & Endy, D. Refactoring bacteriophage T7. *Mol. Syst. Biol.* **1**, 2005.0018 (2005).
3. Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl Acad. Sci. USA* **102**, 15971–15976 (2005).
4. Gibson, D. G. *et al.* Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**, 1215–1220 (2008).
5. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52–56 (2010).
6. Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
7. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–258 (2009).
8. Ro, D.-K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
9. Chin, J. W. Reprogramming the genetic code. *Science* **336**, 428–429 (2012).
10. Mukai, T. *et al.* Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. *Nucleic Acids Res.* **43**, 8111–8122 (2015).
11. Itaya, M., Fujita, K., Ikeuchi, M., Koizumi, M. & Tsuge, K. Stable positional cloning of long continuous DNA in the *Bacillus subtilis* genome vector. *J. Biochem.* **134**, 513–519 (2003).
12. Krishnakumar, R. *et al.* Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. *Nucleic Acids Res.* **42**, e111 (2014).
13. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
14. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6**, 393–403 (2014).
15. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010).
16. Cho, B.-K. *et al.* The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* **27**, 1043–1049 (2009).
17. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
18. Sørensen, M. A. & Pedersen, S. Absolute *in vivo* translation rates of individual codons in *Escherichia coli*. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280 (1991).
19. Curran, J. F. & Yarus, M. Rates of aminoacyl-tRNA selection at 29 sense codons *in vivo*. *J. Mol. Biol.* **209**, 65–77 (1989).
20. Kimchi-Sarfaty, C. *et al.* A “silent” polymorphism in the MDR1 gene changes substrate specificity. *Science* **315**, 525–528 (2007).
21. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
22. Quax, T. E. F. *et al.* Differential translation tunes uneven production of operon-encoded proteins. *Cell Reports* **4**, 938–944 (2013).
23. Quax, T. E. F., Claassens, N. J., Söll, D. & van der Oost, J. Codon bias as a means to fine-tune gene expression. *Mol. Cell* **59**, 149–161 (2015).
24. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* **157**, 624–635 (2014).
25. Pósfai, G. *et al.* Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**, 1044–1046 (2006).

26. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
27. Bryksin, A. V. & Matsumura, I. Rational design of a plasmid origin that replicates efficiently in both gram-positive and gram-negative bacteria. *PLoS One* **5**, e13244 (2010).
28. Kouprina, N., Noskov, V. N. & Larionov, V. Selective isolation of large chromosomal regions by transformation-associated recombination cloning for structural and functional analysis of mammalian genomes. *Methods Mol. Biol.* **349**, 85–101 (2006).
29. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–5035 (1998).
30. Sharp, P. M. & Li, W. H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
31. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–5044 (2004).
32. Tuller, T., Waldman, Y. Y., Kupiec, M. & Ruppin, E. Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl Acad. Sci. USA* **107**, 3645–3650 (2010).
33. Gerdes, S. Y. *et al.* Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684 (2003).
34. Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–D612 (2013).
35. Dai, K. & Lutkenhaus, J. The proper ratio of FtsZ to FtsA is required for cell division to occur in *Escherichia coli*. *J. Bacteriol.* **174**, 6145–6151 (1992).
36. Dewar, S. J., Begg, K. J. & Donachie, W. D. Inhibition of cell division initiation by an imbalance in the ratio of FtsA to FtsZ. *J. Bacteriol.* **174**, 6314–6316 (1992).
37. Lajoie, M. J. *et al.* Probing the limits of genetic recoding in essential genes. *Science* **342**, 361–363 (2013).
38. Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
39. Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016).
40. Napolitano, M. G. *et al.* Emergent rules for codon choice elucidated by editing rare arginine codons in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **113**, E5588–E5597 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements Work was supported by the Medical Research Council, UK (MC_U105181009 and MC_UP_A024_1008, J.W.C.), the Danish Council for Independent Research (DFF – 4090-00289, to J.F.), a Boehringer Ingelheim Fonds PhD fellowship (to S.F.B.), The Gates-Cambridge Scholarship (to S.H.K.), and an ERC Advanced Grant (SGCR to J.W.C.). We thank Neil Grant MRC-LMB Visual Aids for photography.

Author Contributions J.W.C. defined the direction of research. K.W. designed and constructed the REXER and GENESIS systems. J.F. implemented DNA assembly methods in *S. cerevisiae*. K.W. and S.F.B. identified target codons, developed tE and designed recoding schemes. K.W., J.F., S.F.B., S.H.K., and T.C. performed experiments. All authors analysed the data. K.W. and J.W.C. wrote the paper with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W. (kaihang@mrc-lmb.cam.ac.uk) or J.W.C. (chin@mrc-lmb.cam.ac.uk).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Sequences. Sequences of plasmids, BACs, and modified genomic loci are provided in Supplementary Data 1–5.

Construction of selection cassettes, cell strains, and plasmids. Two double selection cassettes were constructed. The $-1/+1$ is a fusion between the negative selection marker *rpsL* (-1) encoding the essential ribosomal protein S12 and conferring sensitivity to streptomycin in *rpsL*K43R genomic background, and the positive selection marker *Kan^R* ($+1$) encoding the kanamycin resistance gene neomycin phosphotransferase II. The *rpsL*–*Kan^R* cassette was expressed as two separate proteins from a single mRNA driven by constitutive transcription from the wild-type *rpsL* promoter. The $-2/+2$ is a fusion between the negative selection marker *sacB* (-2) conferring sensitivity to sucrose, and the positive selection marker *Cm^R* ($+2$) encoding the chloramphenicol resistance gene chloramphenicol acetyl transferase. The *sacB*–*Cm^R* cassette was expressed as two separate proteins from a single mRNA driven by constitutive transcription from the EM7 promoter. Both selection cassettes were synthesized *de novo*.

The minimum genome *E. coli* strain MDS42 was used as the starting strain²⁵. A K43R mutation was introduced into the *rpsL* gene to create MDS42^{*rpsL*K43R} to confer resistance to streptomycin in the absence of an additional wild-type copy of *rpsL* and sensitivity to streptomycin in the presence of any additional copy of wild-type *rpsL*. The $-1/+1$ cassette *rpsL*–*Kan^R* was inserted between positions 89,061 and 89,587 in the MDS42^{*rpsL*K43R} genome to create MDS42^{*rpsL*K43R/rK}.

pKW20_CDFtet_pAraRedCas9_tracrRNA was constructed by assembling multiple PCR fragments using Gibson Assembly. The plasmid backbone and replication origin is from pCDFDuet-1 plasmid (Addgene), in which the spectinomycin resistance marker is replaced with a tetracycline resistance marker from pBR322 plasmid (New England BioLab). The *araC* gene, the arabinose promoter (pAra), and the lambda red (α /beta/ γ) genes were PCR amplified from pRed/ET plasmid (GeneBridges). The open reading frame of Cas9 was PCR amplified from pCas9 plasmid²⁶ and placed downstream of lambda red α . The tracrRNA with its endogenous promoter was PCR amplified from pCas9 plasmid²⁶ and placed in the same orientation downstream of the *araC* gene. The pKW20_CDFtet_pAraRed($\Delta\beta$)Cas9_tracrRNA was derived from pKW20_CDFtet_pAraRedCas9_tracrRNA by inserting GTAC between the 314th and 315th nucleotides of the lambda red beta open reading frame, which leads to translational frame shifting and thus inactivation of lambda red beta.

pKW21_MB1amp_Spacer0 was constructed by assembling two PCR fragments using Gibson Assembly Master Mix (New England BioLab). The pMB1 replication origin and ampicillin resistance marker were PCR amplified from pBR322 plasmid (New England BioLab). The CRISPR array with no functional spacer RNA (hence the nomenclature 0) between BamHI and EcoRI was PCR amplified from pCRISPR²⁶. pKW21_MB1amp_Spacer0 was verified by sequencing. CRISPR arrays with two or four different spacer RNA sequences for directing REXER 2 or REXER 4, respectively, with interspaced direct repeats were commercially synthesized. The arrays were cloned into pKW21_MB1amp_Spacer0, replacing the empty CRISPR array to create different pKW21_MB1amp_Spacers \times 2 or pKW21_MB1amp_Spacers \times 4 plasmids. The final pKW21_MB1amp_Spacers plasmids were sequence verified. A related version of pKW21_MB1erm_Spacers plasmid was prepared by replacing the ampicillin resistance marker in pKW21_MB1amp_Spacers with an erythromycin resistance marker.

The BAC holding the synthetic DNA was constructed by assembling multiple fragments. The BAC backbone is based on pBeloBAC11 (New England BioLabs) from nucleotide 1542 to 7041 with the addition of the double selection cassette $-2/+2$ and the negative selection marker -1 , and assembled using Gibson Assembly Master Mix (New England BioLabs). An alternative arrangement utilizes $-1/+1$ coupled with -2 . The synthetic DNA was always flanked by AvrII sites, which also function as PAMs and part of the protospacer for CRISPR/Cas9.

Assembling short synthetic DNA onto BAC using Gibson Assembly. The pBAC_HR(89,061)–sC-HR(89,587)_r was constructed by assembling three PCR fragments using Gibson Assembly. The first fragment was the 2.2-kb $-2/+2$ *sacB*–*Cm^R* cassette flanked with HR1 (89,012–89,061; all numbering is from the MG1655 reference sequence) and HR2 (89,587–89,636) and further flanked with two AvrII sites; the second fragment was the -1 *rpsL* gene with the *rrnC* terminator; and the third fragment was the pBeloBAC11 backbone from nucleotides 1,542 to 7,041. The assembled pBAC_HR(89,061)–sC-HR(89,587)_r was selected on Luria broth (LB) agar plates with 18 μ g/ml chloramphenicol and sequence verified. The pBAC_HR(89,061)–rK-HR(89,587)_s was similarly constructed using the $-1/+1$ *rpsL*–*Kan^R* cassette flanked with HR1(89,012–89,061) and HR2(89,587–89,636)

and further flanked with two AvrII sites, the -2 *sacB* gene with *rrnC* terminator, and the pBeloBAC11 backbone. The pBAC_HR(89,061)–T5Lux–sC-HR(89,587)_r was constructed by inserting a PCR product of an artificial *lux* operon between HR1 and the $-2/+2$ *sacB*–*Cm^R* cassette in the pBAC_HR(89,061)–sC-HR(89,587)_r.

Assembling long synthetic DNA onto BAC using recombination in *S. cerevisiae*. Long synthetic DNA fragments (≥ 20 kb) were assembled in *S. cerevisiae*. The pBeloBAC11 backbone was converted into a BAC/YAC shuttle vector by introducing a *S. cerevisiae* replication centromere CEN and URA3 selection marker (from *S. cerevisiae* vector pRS316, Addgene). The BAC/YAC shuttle vector holding long synthetic DNA was assembled from 5–16 DNA fragments in *S. cerevisiae*²⁸.

Classical recombination and simultaneous double selection recombination protocol. The *sacB*–*Cm^R* cassette was PCR amplified using primers containing HR1 and HR2. In classical recombination, 3 μ g of this purified PCR product was transformed into 100 μ l electro-competent MDS42^{*rpsL*K43R/rK} cells, which were pre-transformed with the pRed/ET plasmid and induced to express the lambda red components. The cells were recovered in 4 ml super optimal broth (SOB) medium for 1 h at 37 °C and then diluted to 100 ml LB and incubated for 4 h at 37 °C with shaking. The culture was then spun down and re-suspended in 4 ml of LB and spread in serial dilutions on selection plates of LB agar with 18 μ g/ml chloramphenicol.

In simultaneous double selection recombination (DOSER), 3 μ g of the same PCR product was transformed into 100 μ l electro-competent MDS42^{*rpsL*K43R/rK} cells following the same transformation and recovery protocol as above. The culture was then spun down and re-suspended in 4 ml LB and spread in serial dilutions on selection plates of LB agar with 18 μ g/ml chloramphenicol and 50 μ g/ml streptomycin.

Multiple colonies from classical recombination and DOSER were picked for phenotyping. Colony PCRs of multiple clones from classical recombination and DOSER were performed using primer pairs flanking the genomic locus 89,061 to 89,587 with MDS42^{*rpsL*K43R/rK}, MDS42^{*rpsL*K43R}, and Milli-Q filtered water with no resuspended colony as controls. All PCR products were run in parallel to NEB 2-Log DNA Ladder (New England BioLab) and sequence verified by Sanger sequencing.

REXER protocol. MDS42^{*rpsL*K43R/rK} cells were double transformed with pKW20_CDFtet_pAraRedCas9_tracrRNA and pBAC_HR(89,061)–sC-HR(89,587)_r and plated on LB agar plates supplemented with 2% glucose, 10 μ g/ml tetracycline and 18 μ g/ml chloramphenicol. Individual colonies were inoculated into LB medium with 10 μ g/ml tetracycline and 18 μ g/ml chloramphenicol, and grown overnight at 37 °C with shaking. The overnight culture was diluted in LB medium with 10 μ g/ml tetracycline and 18 μ g/ml chloramphenicol to OD₆₀₀ = 0.05 and grown at 37 °C with shaking for around 3 h until OD₆₀₀ \approx 0.3. Arabinose powder was added to the culture to reach a final concentration of 0.5% and the culture was incubated for one additional hour at 37 °C with shaking. The cells were harvested at OD₆₀₀ \approx 0.6, and made electro-competent in 1/500th of the culture volume.

We electroporated 3 μ g pKW21_MB1amp_Spacers \times 2 or pKW21_MB1amp_Spacers \times 4 plasmid into 100 μ l of competent cells. The cells were recovered in 4 ml SOB medium for 1 h at 37 °C and then diluted to 100 ml LB supplemented with 50 μ g/ml ampicillin and 10 μ g/ml tetracycline and incubated for 4 h at 37 °C with shaking. The culture was spun down and re-suspended in 4 ml LB and spread in serial dilutions on selection plates of LB agar with 18 μ g/ml chloramphenicol and 50 μ g/ml streptomycin. The plates were incubated at 37 °C overnight, and the efficiency was calculated by counting colonies. Multiple colonies were picked, resuspended in Milli-Q filtered water, and arrayed on LB agar plates or LB agar plates supplemented with 18 μ g/ml chloramphenicol or 50 μ g/ml kanamycin. Colony PCR was also performed from resuspended colonies using the primer pair flanking the genomic locus 89,061–89,587.

The resulting colonies with the $-2/+2$ *sacB*–*Cm^R* cassette replacing the $-1/+1$ *rpsL*–*Kan^R* cassette at the genomic locus 89,062–89,586 were incubated in LB without ampicillin, to lose the pKW21_MB1amp_Spacers \times 2 or pKW21_MB1amp_Spacers \times 4 plasmid. The resulting cells were double transformed with pKW20_pCDFtet_pAraRedCas9_tracrRNA and pBAC_HR(89,061)–rK-HR(89,587)_s. An individual colony was picked, inoculated into LB, and prepared into electro-competent cell as previously described. 3 μ g pKW21_MB1erm_Spacers \times 2 or pKW21_MB1erm_Spacers \times 4 plasmid was electroporated into the pre-induced cell. The cells were recovered in 4 ml SOB medium for 1 h at 37 °C and then diluted to 100 ml LB supplemented with 25 μ g/ml erythromycin and 5 μ g/ml tetracycline and incubated for 4 h at 37 °C with shaking. The culture was spun down and re-suspended in 4 ml LB and spread in serial dilutions on selection plates of LB agar with 3% sucrose and 25 μ g/ml kanamycin. After incubating the selection plate at 37 °C overnight, multiple colonies were picked, resuspended in Milli-Q filtered water, and arrayed on LB agar plates, or LB agar plates

supplemented with 18 µg/ml chloramphenicol or 50 µg/ml kanamycin. Colony PCR was performed from resuspended colonies using the primer pair flanking the genomic locus 89,061–89,587.

The pBAC_HR(89,061)-T5Lux-sC-HR(89,587)_r, pBAC_HR(89,061)-90kb/Lux-sC-HR(89,587)_r, pBAC_HR(89,061)-100kb/Lux-sC-HR(192,744)_r, and pBAC_HR(89,061)-20kb-sC-HR(106,508)_r with matching pKW21_MB1amp_Spacers × 2 or pKW21_MB1amp_Spacers × 4 plasmids were used in the other REXER experiments following the same protocol. Colony PCR of the *lux* operon and the coupled $-2/+2$ *sacB*-*Cm^R* cassette inserted at the genomic locus 89,061–89,587 using the primer pair flanking the genomic locus generated a 9-kb band for successful insertion and a 1.5-kb band for the MDS42^{rk} control. Primer pairs flanking the 5' or 3' end of the inserted or replaced DNA were used, which generated a PCR band for correct insertion or replacement, and no band or a band of the wrong size with the MDS42^{rk} control. Colony PCR using primers for the internal watermarks was also performed. The 20-kb recoded region (from 89,062 to 106,507) was PCR amplified from purified genomic DNA (QIAGEN DNeasy Blood & Tissue Kit) using a primer pair flanking the whole region. The 20-kb PCR product was purified using Bio-Rad PCR Kleen Columns and fully sequenced by Sanger sequencing.

Choice of region for systematic and defined synonymous recoding. We applied a sliding window approach, in which we counted the number of target codons in all essential genes within a 10-kb region of the MDS42 genome. Starting from the first 10 kb of the genome sequence, we iteratively shifted the window by 100 nt and performed the codon analysis until the end of the MDS42 genome sequence. Gene essentiality was defined by transposon insertion densities from a genome-scale genetic footprinting study in *E. coli*³³, which led to comparable results to those obtained when we used the KEIO collection data⁴¹.

Choice of recoding rules. We characterized all serine, leucine, and alanine codons using the codon adaptation index (cAi)³⁰ and the tRNA adaptation index (tAi)^{31,32}. In the case of cAi, we used the relative adaptiveness of each codon *i* (expressed as cAi_{*i*}) as a metric. In the case of tAi, we used the relative adaptiveness value of each codon *i* (expressed as tAi_{*i*}) in Supplementary Table 2^{31,32}. We defined ideal synonymous substitutions for targeted codons by minimising the difference between cAi_{*i*} and tAi_{*i*}. Comparing cAi_{*i*} and tAi_{*i*} for all codons, we noticed that the two metrics did not correlate well (Pearson's $R^2 = 0.24$) and decided to propose a third metric. In particular, we assumed that translation efficiency increases proportionally with increasing isoacceptor tRNA concentration and decreases proportionally with increasing numbers of competing codons that are translated by the same isoacceptor tRNA. On this basis we defined the translation efficiency (TE) of codon *i* as follows:

$$tE_i = \sum_j \left(\frac{k_{ij} \times [tRNA_j]}{\sum_m q_m k_{mj}} \right), k_{ij}, k_{mj} \in \left\{ \begin{array}{l} \text{cognate: 1.0} \\ G - U/U - G \text{wobble: 0.5} \\ C/U - xo^5U: 0.25 \\ C/U - \text{inosine: 0.1} \\ A - \text{inosine: 0.05} \end{array} \right\},$$

where codon *i* is translated by tRNAs *j*, *k_{ij}* denotes the interaction strength between codon *i* and tRNA *j*, *m* denotes each codon translated by tRNA *j*, and *k_{mj}* denotes the interaction strength between codon *m* and tRNA *j*. The interaction strengths were defined in five groups: i) 'cognate' for codons that are reverse complements to the respective tRNA anticodon as well as AUA^{lle}-k²CAU^{tRNA}; ii) 'G–U/U–G wobble' for codons where a third position G or U interacts with a (modified) tRNA U or G, respectively; iii) 'C/U – xo⁵U' for codons where a third position C or U interacts with an xo⁵-modified uridine in the tRNA anticodon; iv) 'C/U – inosine' where a third position C or U in the codon interacts with an inosine in the tRNA anticodon (an interaction shown to be 3–8-fold weaker than G–U wobbling⁴²); and v) 'A – inosine' for the reportedly weak interaction between the third position A in a codon with an inosine in the tRNA anticodon⁴³. We obtained the tRNA concentrations [tRNA_{*j*}] from reported measurements performed on *E. coli* cultures, expressed as a fraction of tRNA out of total tRNA in per cent⁴⁴. To determine the relative transcriptomic codon frequency *q* for each codon *i* we first calculated the codon's absolute transcriptomic frequency *r_i*:

$$r_i = \sum_x g_{ix} \times t_x,$$

where *g_{ix}* is the frequency of codon *i* in gene *x* and *t_x* is the transcript abundance of gene *x* according to empirical data (DNA array data for wild-type *E. coli* grown at 0.5 h^{–1})⁴⁵. Finally *r_i* was transformed into *q_i* by dividing *r_i* by the maximal value found for *r* across all codons:

$$q_i = \frac{r_i}{\max(r)}.$$

Using the three coding metrics, we constructed Extended Data Table 1 by assigning the closest substitutions (in pink) for synonymous recoding of TCA^{Ser} and TCG^{Ser} (in grey).

Individual recoding landscapes and compiled recoding landscapes. Based on the complete sequence of the recoded region for each clone, the identity of the codon at each of the attempted recoding positions and the duplicated region 1 to 5 was identified either as recoded, with a binary value of 1 and coloured red, or wild type, with a value of 0 and coloured black. The distribution of targeted positions that were recoded and that remained wild type across the refactored *mraZ* to *ftsZ* region gives an 'individual recoding landscape'.

The 16 individual recoding landscapes of the 16 individual clones of each of the recoding schemes were compiled to generate the 'compiled recoding landscape' of each recoding scheme by counting the fraction of clones being recoded at each targeted position across the whole refactored *mraZ* to *ftsZ* region. When the recoding fraction at a given position is greater than 0 (red), it indicates that there is at least one sequenced clone being recoded at this position. When the recoding fraction reaches 0 (black), it indicates that the wild-type codon always remains and that the recoded codon may not be tolerated at these positions.

REXER + ssDNA recombineering protocol. A single-stranded oligo of a total length of 90 nt was designed and synthesized to change the deleterious sequence of AGT in *ftsA* codon position 407 on the synthetic sequence of r.s.1 to a fixing sequence of AGC. The oligo sequence was designed based on the reverse strand of the synthetic sequence to bind the forward strand with the single nucleotide change positioned in the middle (position 45 from 5' end). The last two nucleotides in the 5' end of the oligo were substituted with a phosphorothioate backbone to protect the oligo from unspecific exonuclease degradation.

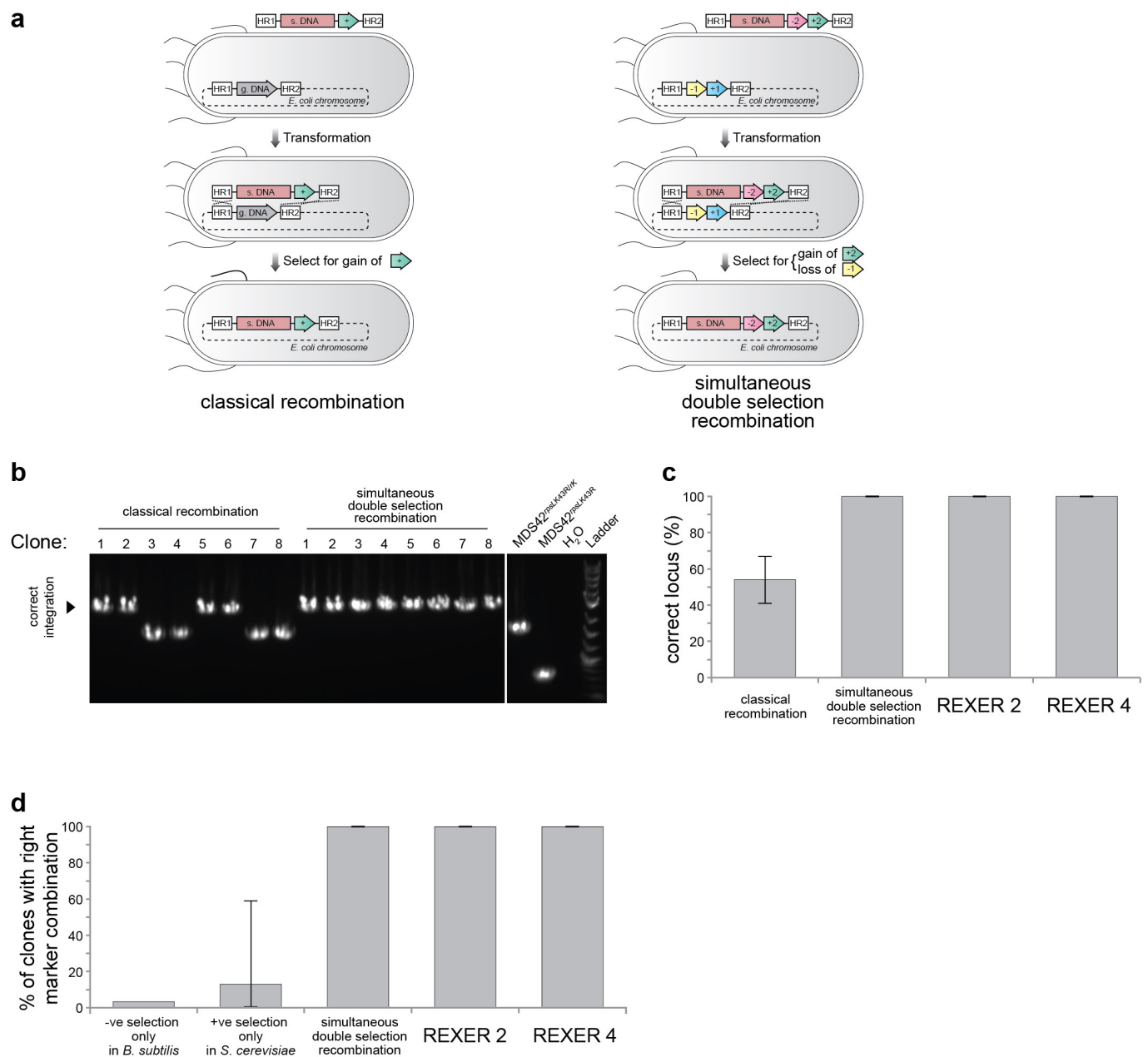
We co-electroporated 3 µg of matching pKW21_MB1amp_Spacers × 2 plasmid and 0.2 nmol of the fixing oligo into the pre-induced MDS42^{tpsLK43R/rK} cells with pCDFtet_pAraRedCas9_tracrRNA and pBAC_HR(89,061)-20kb/r.s.1-sC-HR(106,508)_r. The normal REXER procedure was carried out without any modification. SNP-PCRs from re-suspended survival colonies were performed⁴⁶ using the SNP-PCR primer pairs specific for either the *ftsA* codon position 407 wild-type sequence TCG or the fixed sequence AGC⁴⁷ with KAPA 2G fast multiplex mix, and analysed on QIAGEN QIAxcel Advanced using QIAxcel DNA Screening Kit with QX Alignment Marker 15 bp/5 kb and QX Size Marker 250–4000 bp. Clones with the correct genotype following REXER + ssDNA recombineering were verified by Sanger sequencing through the entire 20-kb region.

Growth rate measurements and analysis. Glycerol stocks of the assayed bacterial clones were used to inoculate 5 ml LB in the absence of antibiotics for overnight incubation at 37 °C with shaking. The overnight cultures were used to inoculate triplicates of 1 ml LB in a deep-well pre-culture plate at a ratio of 1:100, followed by incubation at 37 °C for 6 h with shaking. Each replicate on the pre-culture plate was used to inoculate 200 µl LB in a 96-well measurement plate with a dilution factor of 100. The measurement plate was incubated at 37 °C for 16 h with shaking at 400 rpm in an M200 Pro Plate Reader (Tecan). Readings of OD₆₀₀ were taken for each well every 10 min. Plate reader absorbance data was adjusted to correspond to spectrophotometer readings by collecting measurements from a dilution series of bacterial cultures and fitting the plate reader data *y* with a polynomial to the spectrophotometer values *x*: $y = 2.053x^2 + 2.2x + 0.061$.

To determine doubling times, the growth curves were log₂-transformed, and the first derivative was determined (d(log₂(*x*))/dt). During exponential growth, the log₂-derivative is maximal and constant. The ten time-points with the maximal log₂-derivatives were identified and used to calculate the average value with standard deviation for each of the replicates. A total of 12 replicates (three independent clones, each independently repeated four times) were used to calculate the doubling time for each fully recoded scheme. For each doubling time, the average across the *n* = 12 replicates was determined and the error σ was propagated using the following formula: $\sqrt{\sum_{i=1}^n \sigma^2 / \sqrt{n}}$.

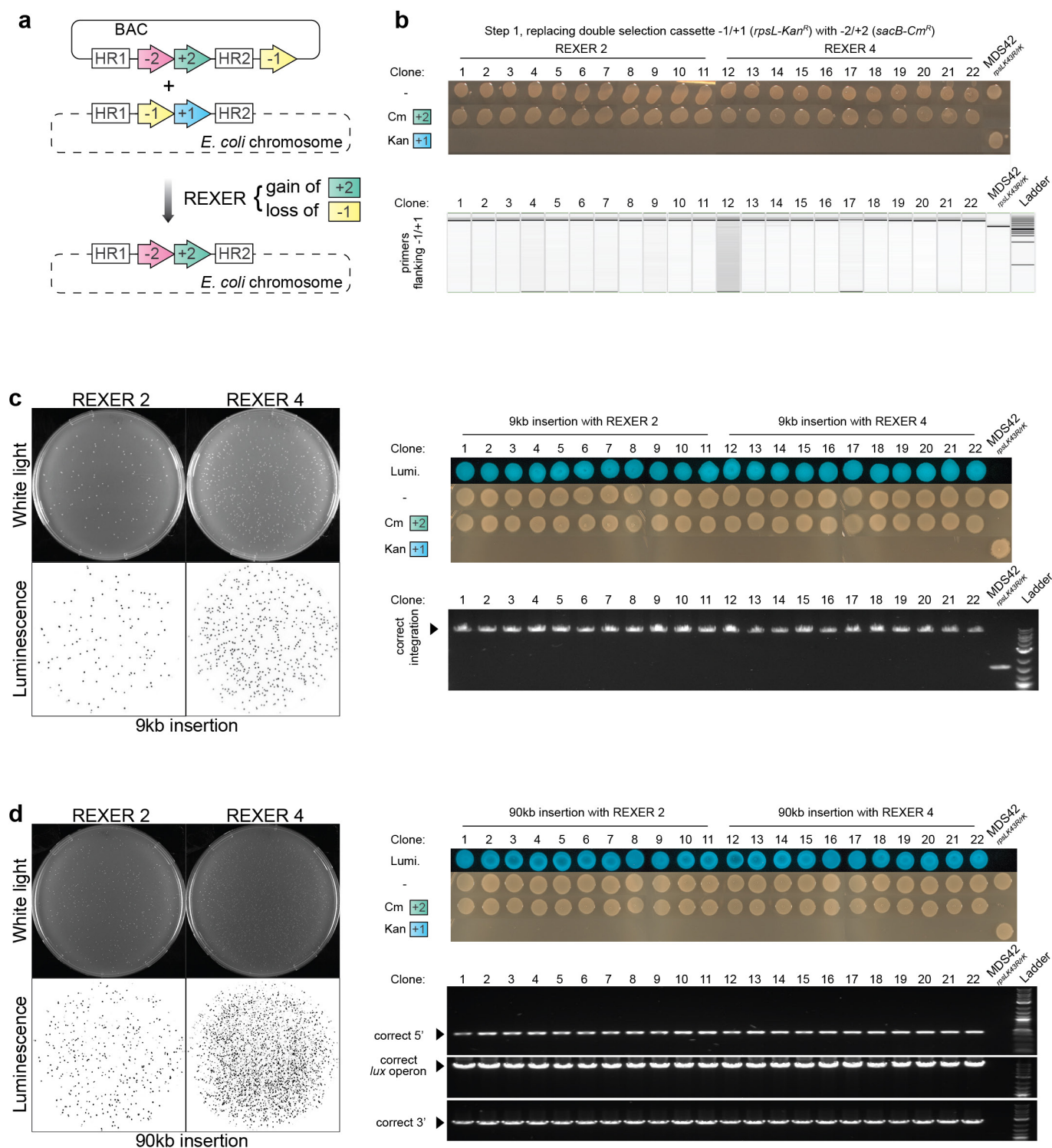
Data availability. The sequences used in this study are available in Supplementary Data 1–5. Supplementary Data 1: Genomic locus for selection marker –1/+1; Supplementary Data 2: Plasmid containing lambda red, Cas9, and tracrRNA; Supplementary Data 3: BAC containing *lux* operon and –2/+2 for integration; Supplementary Data 4: Plasmid containing spacers for REXER 2; Supplementary Data 5: Plasmid containing spacers for REXER 4. All other datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

41. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
42. Grosjean, H. J., de Henau, S. & Crothers, D. M. On the physical basis for ambiguity in genetic coding interactions. *Proc. Natl Acad. Sci. USA* **75**, 610–614 (1978).
43. Curran, J. F. Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res.* **23**, 683–688 (1995).
44. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J. Mol. Biol.* **260**, 649–663 (1996).
45. Ishii, N. *et al.* Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* **316**, 593–597 (2007).
46. Gallagher, R. R., Li, Z., Lewis, A. O. & Isaacs, F. J. Rapid editing and evolution of bacterial genomes using libraries of synthetic DNA. *Nat. Protocols* **9**, 2301–2316 (2014).
47. Newton, C. R. *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2516 (1989).



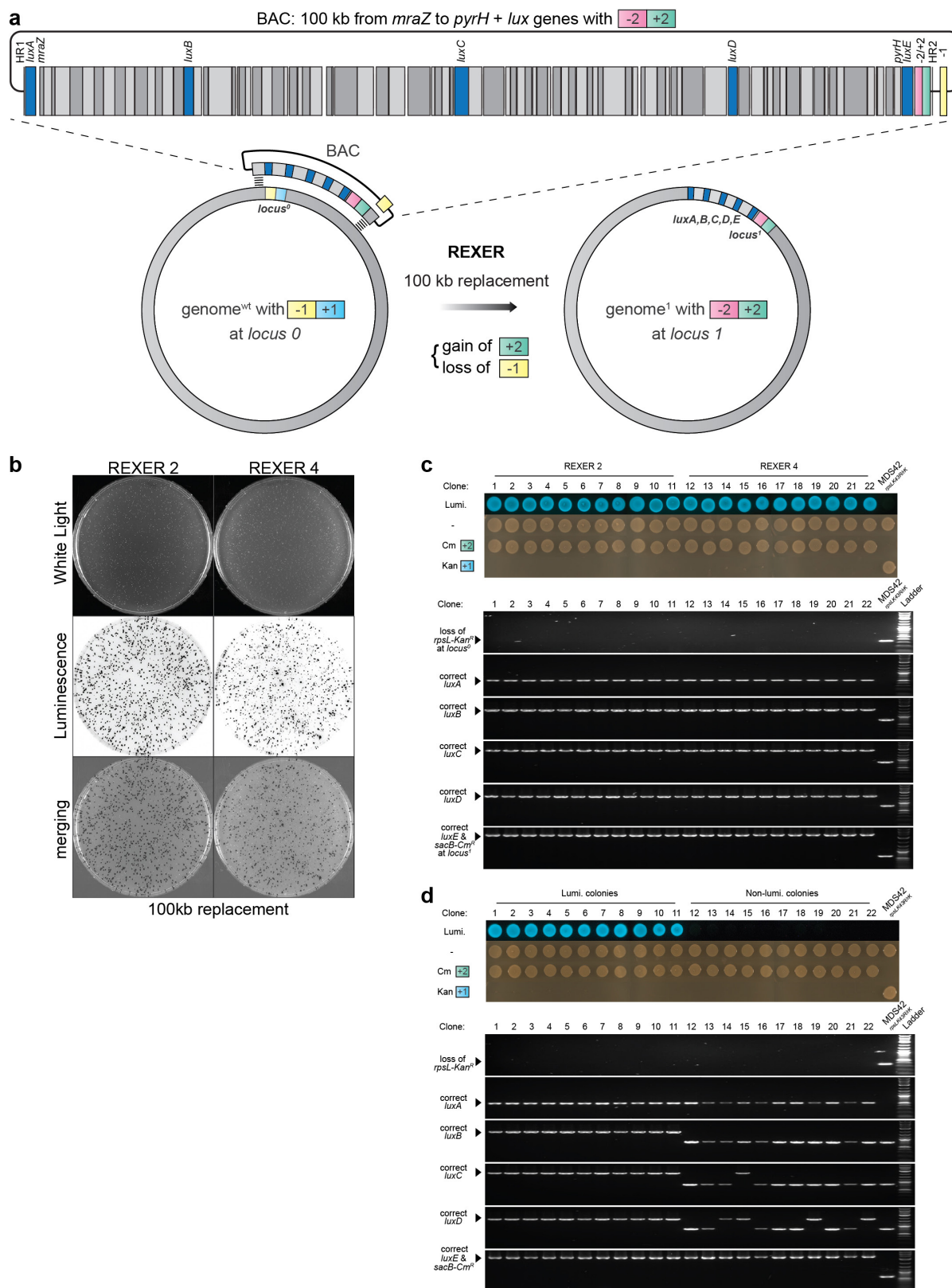
Extended Data Figure 1 | Simultaneous double selection and recombination enhances integration at a target locus. **a**, Classical recombination and double selection recombination. In classical recombination, a linear dsDNA with a synthetic DNA (s. DNA) sequence and a positive selection marker (+, Cm^R) flanked by homologous region 1 (HR1) and homologous region 2 (HR2) is transformed into the cell. Recombinants are selected by expression of the positive selection marker. By simultaneous double selection recombination, s. DNA containing the double selection marker $-2/+2$ ($sacB-Cm^R$) is integrated in place of the double selection marker $-1/+1$ ($rpsL-Kan^R$) on the genome. Double selection for the gain of $+2$ and loss of -1 selects for simultaneous gain of s. DNA and loss of genomic sequence, and improves recombination at the target genomic locus. **b**, Colony PCR of clones from classical recombination and simultaneous double selection and recombination. **c**, All of the clones isolated by simultaneous double selection and recombination have s. DNA integrated at the target locus. The data show the mean percentage \pm s.d. at the correct locus ($n = 6$, three biological replicates each performed in two technical replicates, for each technical replicate 8 clones were phenotyped). $P < 10^{-4}$, two tailed t -test for the null

hypothesis that classical recombination is as efficient as double selection recombination, REXER 2 or REXER 4. **d**, The data for simultaneous double selection recombination, REXER 2 and REXER 4 show the mean percentage \pm s.d. at the correct locus ($n = 6$, three biological replicates each performed in two technical replicates, for each technical replicate eight independent clones were phenotyped). The data for *B. subtilis* and *S. cerevisiae* are from previous publications. A previously reported method integrating foreign DNA into *B. subtilis* genome only using negative selection gave 3% (9 out of 271) of selected clones with right combination of markers^{3,11}. A previously reported method replaced the *S. cerevisiae* chromosome III in 11 steps using only positive selection. The efficiency, as judged by clones with the correct combination of markers, was reported for ten of these steps; the mean percentage of clones with the right combination of markers is plotted (13%). The error bar represents the maximum and minimum integration efficiency as judged by clones with the correct combination of markers. The minimum efficiency was 0.5% (replacement of 55 kb), the maximum efficiency was 59% (replacement of 9 kb)⁶. For gel source images, see Supplementary Fig. 1.



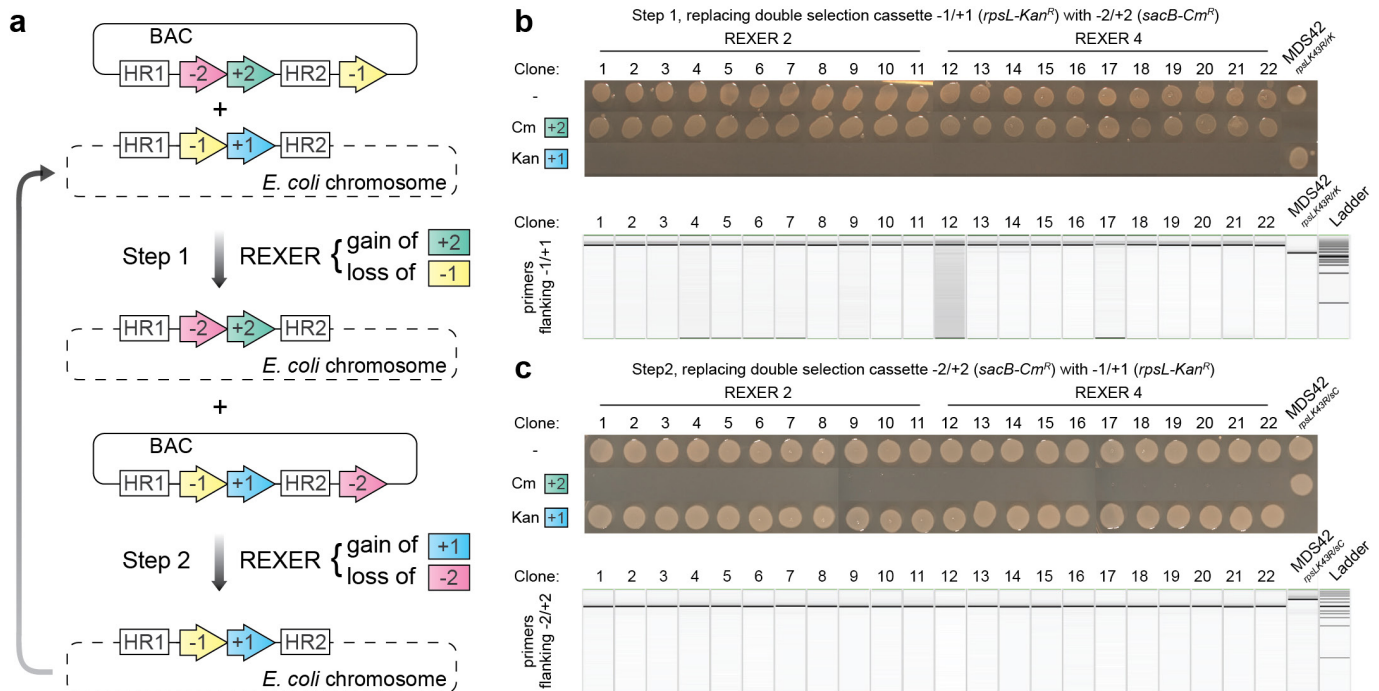
Extended Data Figure 2 | REXER enables site-specific integration of large DNA fragments into the genome. **a**, The use of two distinct double selection cassettes ($-1/+1$ (*rpsL-Kan^R*) and $-2/+2$ (*sacB-Cm^R*) allows simultaneous selection for the loss of the negative selection marker on the genome and the gain of the positive selection marker from the BAC upon integration of synthetic DNA. **b**, Efficient replacement of genomic *rpsL-Kan^R* with BAC-bound *sacB-Cm^R* using REXER 2 and REXER 4. All colonies tested ($n = 22$) contained the correct combination of selection markers after REXER 2 or REXER 4 as analysed by phenotyping, colony PCR, and DNA sequencing (not shown). **c**, Efficient insertion of 9-kb

synthetic DNA. Genomic *rpsL-Kan^R* was replaced with a synthetic *lux* operon coupled to *sacB-Cm^R* using REXER 2 and REXER 4. All colonies on the tenfold dilution double selection plates for REXER 2 and the 10^4 -fold plates for REXER 4 show bioluminescence. Eleven colonies each from REXER 2 and REXER 4 showed correct integration by phenotyping, colony PCR, and DNA sequencing (not shown). **d**, Efficient insertion of 90-kb synthetic DNA. The 90-kb DNA consisted of the *lux* operon in the middle of 80-kb DNA (previously deleted from the MDS42 genome) and followed by *sacB-Cm^R*, carried on a BAC. For gel source images, see Supplementary Fig. 1.

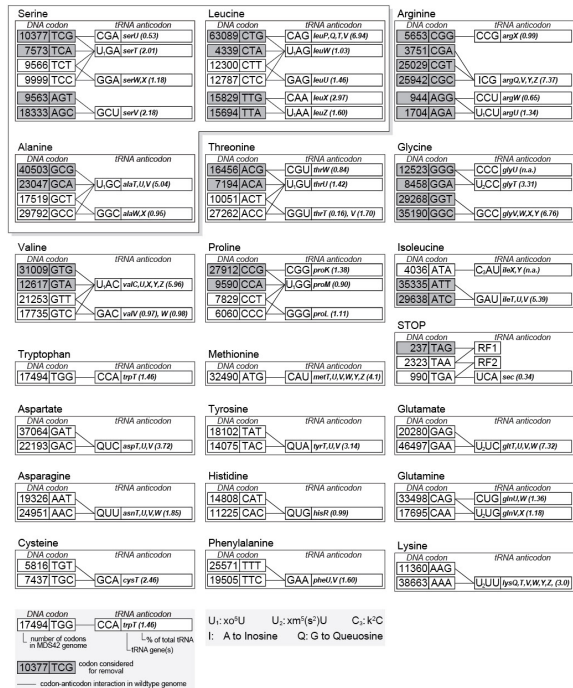
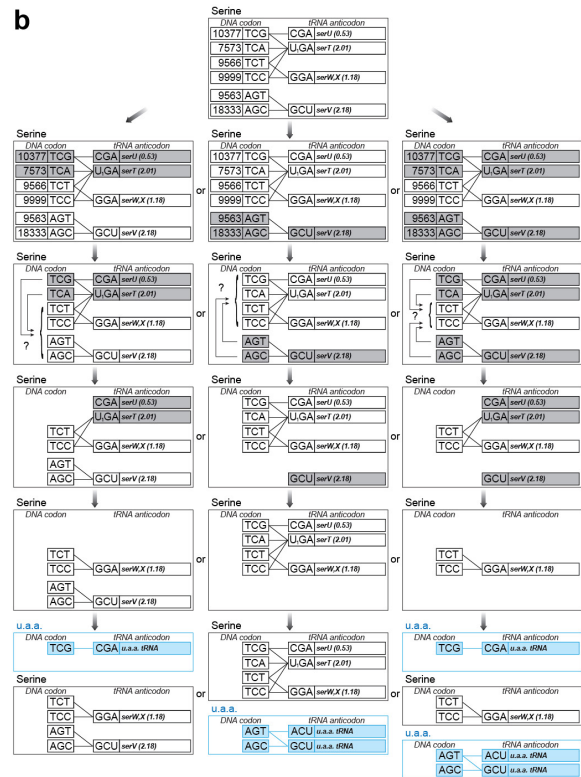
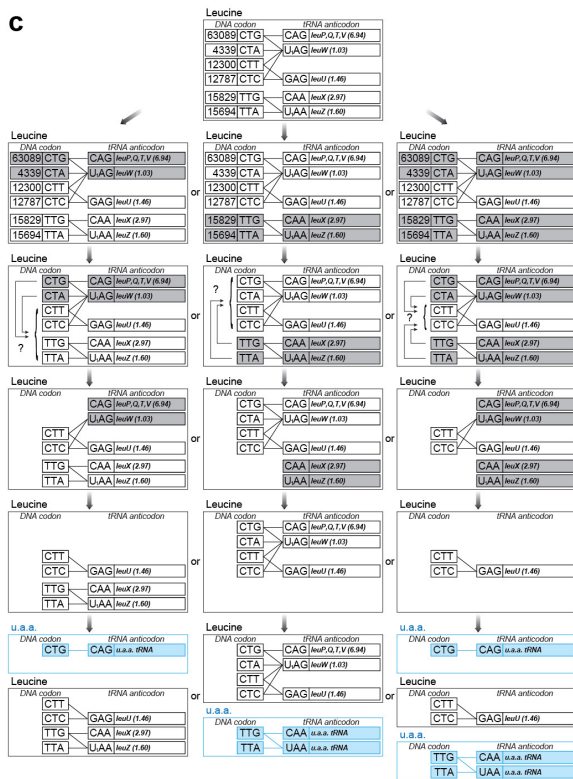
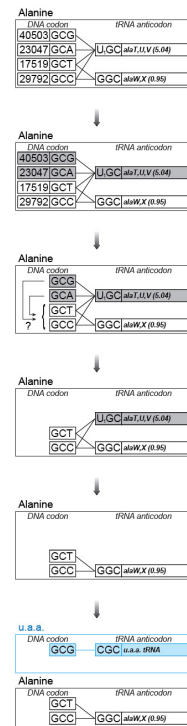


Extended Data Figure 3 | Replacement of 100 kb of genomic DNA via REXER. **a**, The synthetic DNA contains the 100-kb wild-type DNA (open reading frames in grey) with five genes of the *lux* operon (blue) and *sacB-Cm^R*. Complete replacement leads to integration of all five *lux* genes (*luxA*, *B*, *C*, *D* and *E*) resulting in bioluminescent cells, while partial replacement confers loss of one or more *lux* genes and loss of bioluminescence. **b**, After REXER 2, 80% of 2×10^2 colonies examined were bioluminescent; after REXER 4, 50% of 2×10^2 colonies examined were bioluminescent. **c**, Eleven bioluminescent colonies from REXER 2 and eleven bioluminescent colonies from REXER 4 were analysed.

All colonies analysed had all five *lux* genes correctly integrated, indicating complete replacement of the 100-kb genomic region. All clones analysed contained the right combination of selection markers. **d**, Eleven bioluminescent colonies from REXER 2 and eleven non-bioluminescent colonies from REXER 4 were analysed. While bioluminescent colonies contained all five *lux* watermarks, all the non-bioluminescent colonies analysed were lacking one or more *lux* genes, indicating partial replacement of the genomic region. All clones analysed contained the right combination of selection markers. For gel source images, see Supplementary Fig. 1.



Extended Data Figure 4 | Iterative REXER. **a**, The product of REXER shown in Extended Data Fig. 2a was used as a template for the next round of REXER. **b**, The phenotypes of clones from the first round of REXER. **c**, The phenotypes of clones from the second round of REXER. For gel source images, see Supplementary Fig. 1.

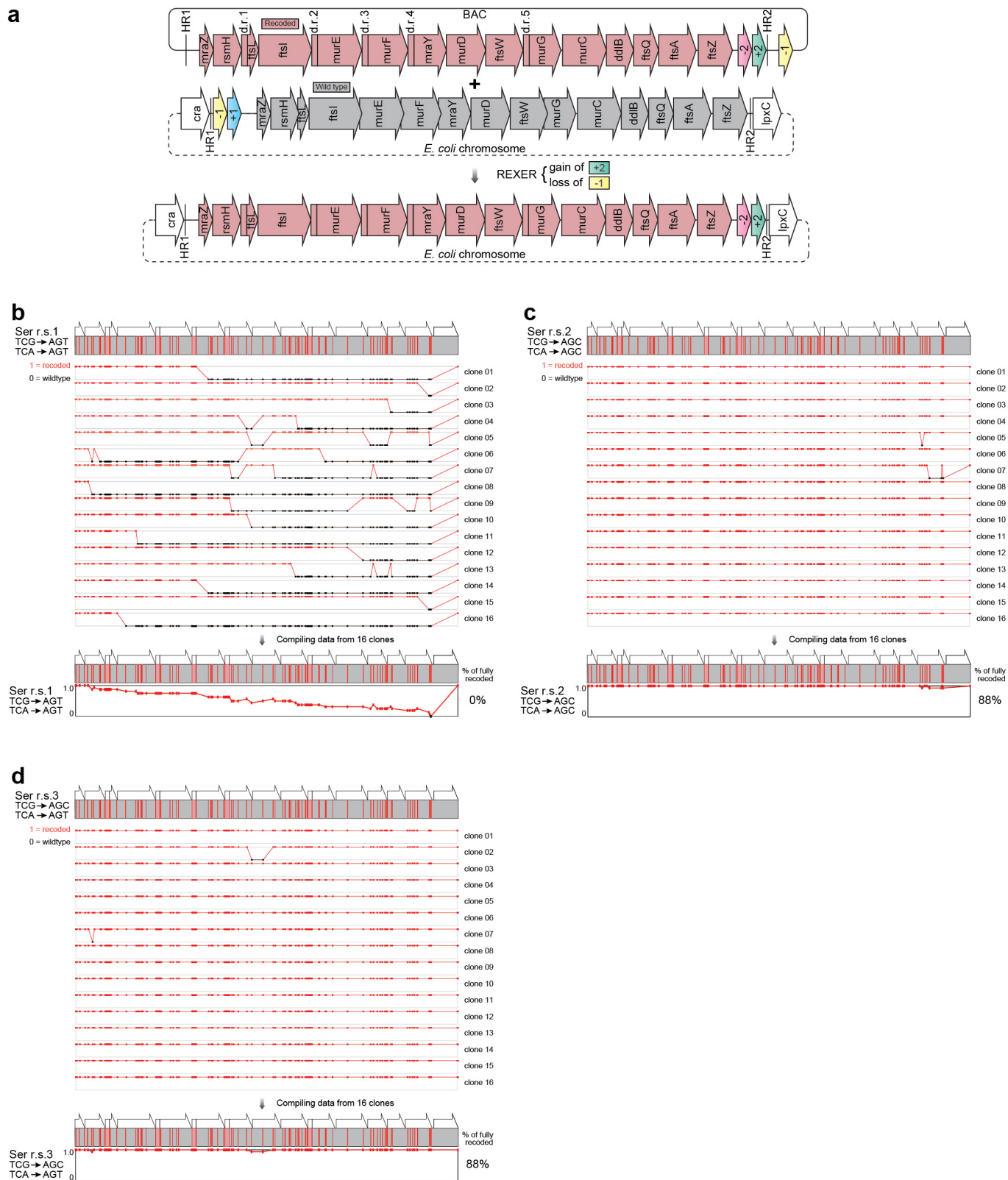
a**b****c****d**

Extended Data Figure 5 | See next page for caption.

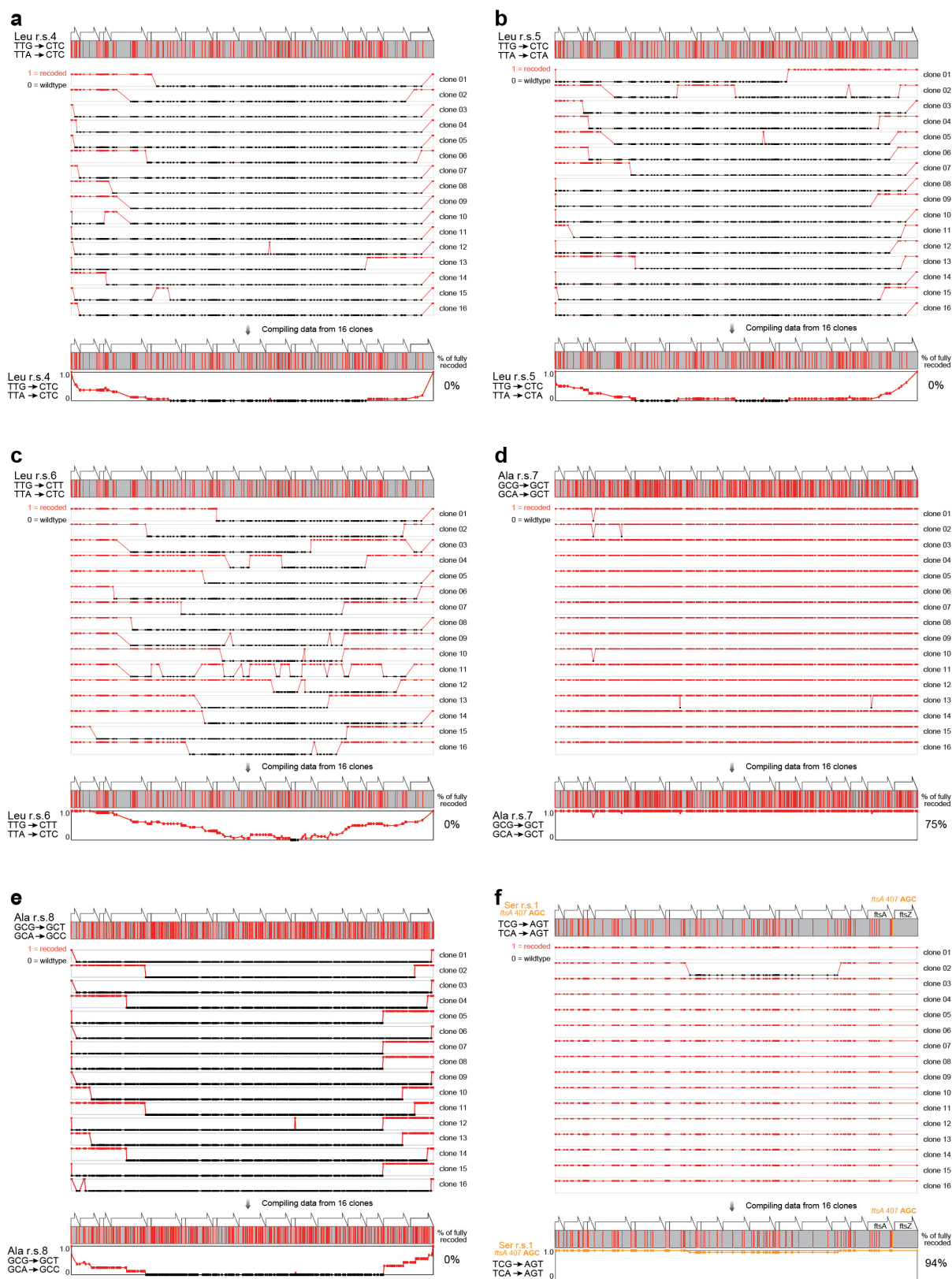
Extended Data Figure 5 | Synonymous codon compression strategies.

a, Codon and anticodon interactions in the *E. coli* genome. Twenty-eight sense codons are highlighted in grey, along with the amber stop codon. The genome-wide removal of these sense codons, but not other sense codons, would enable all their cognate tRNA to be deleted without removing the ability to decode one or more sense codons remaining in the genome. This is necessary but not sufficient for the reassignment of sense codons to unnatural monomers. Serine, leucine and alanine codon boxes are highlighted because the endogenous aminoacyl-tRNA synthetases for these amino acids do not recognize the anticodons of their cognate tRNAs. This may facilitate the assignment of codons within these boxes to new amino acids through the introduction of tRNAs bearing cognate anticodons that do not direct mis-aminacylation by endogenous

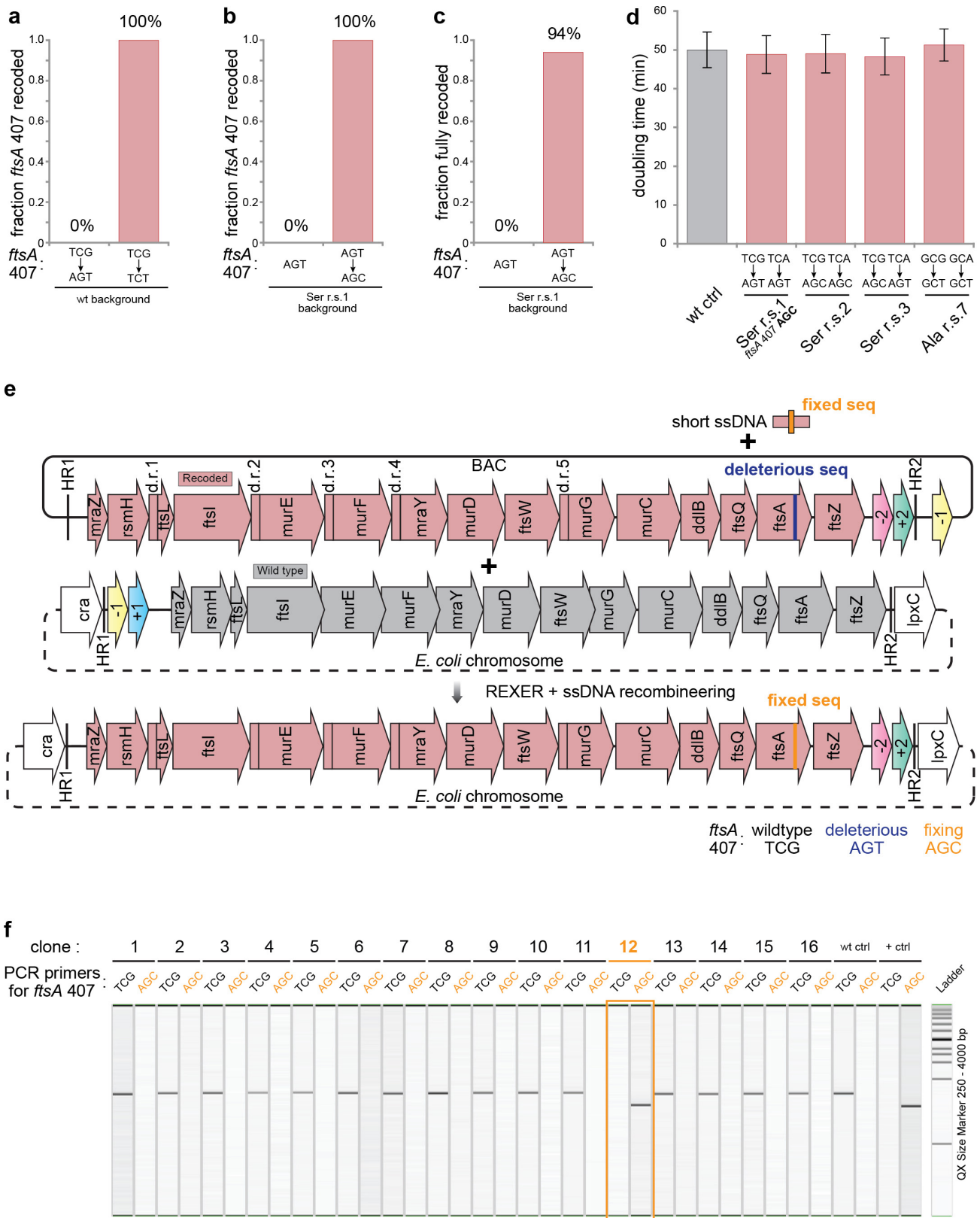
synthetases. The number of total codon counts for all 64 triplet codons in the MDS42 genome (GenBank accession number AP012306), all known codon–anticodon interactions through both Watson–Crick base-pairing and wobbling, base modification on tRNA anticodons, tRNA genes, and measured *in vivo* tRNA relative abundance are reported. This analysis identifies 10 codons from the serine, leucine, and alanine groups (serine codon TCG, TCA, AGT, AGC; leucine codon CTG, CTA, TTG, TTA; and alanine codon GCG, GCA) that satisfy both the codon–anticodon interaction and aminoacyl-tRNA synthetases recognition criteria for codon reassignment. **b–d**, Serine, leucine and alanine codon removal and tRNA deletion strategies compatible with codon reassignment to unnatural amino acids (u.a.a.).



Extended Data Figure 6 | Recoding landscapes for compression of serine codons by REXER. **a**, The sequences for the systematically recoded *mraZ* to *ftsZ* region were *de novo* designed, synthesized and assembled into BAC and used for REXER. **b–d**, The recoding landscapes for serine recoding schemes (r.s.) 1–3, and the resulting compiled recoding landscape.



Extended Data Figure 7 | Recoding landscapes. a–e, Recoding schemes 4–8. **f,** Recoding scheme 1 with *ftsA* codon 407 changed from AGT to AGC (highlighted in orange).



Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | Identifying and fixing a deleterious sequence in defined and systematic synonymous recoding. **a**, Recoding codon 407 in *ftsA* in the wild-type genomic background. The wild-type codon at *ftsA* codon position 407 is the serine codon TCG. We sequenced 16 post-REXER clones for TCG to AGT and 20 post-REXER clones for TCG to TCT. **b**, Changing *ftsA* 407 AGT to AGC in the serine r.s.1 background. We sequenced 16 AGT clones and 16 AGT to AGC clones. **c**, Changing *ftsA* 407 AGT to AGC in the serine r.s.1 background greatly improved the fraction of fully recoded clones across the entire 20-kb region from 0% to 94% (16 clones sequenced). **d**, The fixed serine r.s.1 with *ftsA* 407 AGC yielded clones with no measurable growth defect. The doubling times of fully recoded clones from serine r.s.1 with *ftsA* 407 AGC, serine r.s.2, serine r.s.3, and alanine r.s.7 were measured and showed no measurable growth defects when compared to the wild-type MDS42 *E. coli* control with the second double selection cassette integrated at the same genomic locus. (The *P* values for the null hypothesis that the doubling times of each recoded clone is different from the wild-type control were calculated by two-tailed *t*-tests. Serine r.s.1 *ftsA* 407 AGC versus wild type, *P* = 0.54; serine r.s.2 versus wild type, *P* = 0.62; serine r.s.3 versus wild type, *P* = 0.39; alanine r.s.1 versus wild type, *P* = 0.47.) *n* = 12 biological replicates and error bars show s.d. **e**, Combining single-strand DNA

recombineering with REXER to fix a short deleterious stretch within the synthetic sequence of r.s. 1. A 90-nt single-stranded oligonucleotide was designed to change the deleterious sequence of AGT in *ftsA* codon position 407 in r.s.1 to a tolerated sequence, AGC. The oligonucleotide sequence was designed based on the reverse strand of the synthetic sequence to bind the forward strand with the single nucleotide change positioned in the middle (45 from nt 5' end). The oligonucleotide was co-transformed into *E. coli* during a REXER experiment that introduced r.s. 1 into the genome. **f**, Fixing a short deleterious sequence on synthetic DNA with REXER + ssDNA recombineering. Sixteen clones from REXER double selection (described in **e**) were randomly picked and subjected to single nucleotide polymorphism (SNP) genotyping using primers specific for either the wild-type sequence in *ftsA* codon position 407 (TCG) or the fixed sequence (AGC). MDS42^{*rpsLK43R/rK*} was used as the wild-type control and a fully recoded clone from serine r.s.3 with verified *ftsA* 407 AGC as the positive control. SNP genotyping at *ftsA* codon position 407 identified one clone (clone 12, highlighted in orange) out of a total of 16 clones tested with fixed sequence AGC, which was then fully sequenced across the entire 20-kb recoding region and confirmed as fully recoded at all 83 targeted codon positions. For gel source images, see Supplementary Fig. 1.

Extended Data Table 1 | Defining recoding rules by codon adaptation index (cAi), tRNA adaptation index (tAi), and translation efficiency (tE)

a

Codon	cAi		tAi		t.E	
	Metric	Substitution	Metric	Substitution	Metric	Substitution
TCG ^{Ser}	0.017	AGT ^{Ser}	0.165	AGC ^{Ser}	0.086	AGC ^{Ser}
TCA ^{Ser}	0.077	AGT ^{Ser}	0.125	AGC ^{Ser}	0.049	AGT ^{Ser}
TCT ^{Ser}	1.000		0.110		0.034	
TCC ^{Ser}	0.744		0.250		0.057	
AGT ^{Ser}	0.085		0.055		0.044	
AGC ^{Ser}	0.410		0.125		0.088	

b

Codon	cAi		tAi		t.E	
	Metric	Substitution	Metric	Substitution	Metric	Substitution
CTG ^{Leu}	1		0.540		0.098	
CTA ^{Leu}	0.007		0.125		0.010	
CTT ^{Leu}	0.042		0.055		0.036	
CTC ^{Leu}	0.037		0.125		0.069	
TTG ^{Leu}	0.02	CTC ^{Leu}	0.165	CTC ^{Leu}	0.034	CTT ^{Leu}
TTA ^{Leu}	0.02	CTC ^{Leu}	0.125	CTA ^{Leu}	0.068	CTC ^{Leu}

c

Codon	cAi		tAi		t.E	
	Metric	Substitution	Metric	Substitution	Metric	Substitution
GCG ^{Ala}	0.424	GCT ^{Ala}	0.120	GCT ^{Ala}	0.020	GCT ^{Ala}
GCA ^{Ala}	0.586	GCT ^{Ala}	0.375	GCC ^{Ala}	0.040	GCC ^{Ala}
GCT ^{Ala}	1		0.110		0.019	
GCC ^{Ala}	0.122		0.250		0.027	

We defined the best synonymous replacements for the target serine (**a**), leucine (**b**), and alanine codons (**c**) by identifying the closest match for the target codons, as judged by codon adaptation index (cAi), tRNA adaptation index (tAi), or a third metric that combines codon abundance and measured tRNA concentrations to estimate translation efficiency (tE) (see Methods). The table assigns the closest substitutions (in pink) for synonymous recoding of targeted codons (in grey) using the three coding metrics. Where two substitutions are comparable the one that conserves G, C content is chosen. The number in bold is the value of the best matching substitution in a given coding metric.

Extended Data Table 2 | Properties of genes targeted for recoding

a

Gene	Function	Localisation	Protein level ppm	ORF length	Peptide length	Number of target codons							
						r.s.1	r.s.2	r.s.3	r.s.4	r.s.5	r.s.6	r.s.7	r.s.8
mraZ	Transcription factor	cytosol, nucleoid	11.3	459	153	4	4	4	9	9	9	4	4
rsmH	Methyltransferase	cytosol	122.0	942	314	8	8	8	4	4	4	13	13
ftsL	Cell division	membrane	1.9	366	122	2	2	2	5	5	5	4	4
ftsI	Cell division	membrane	9.7	1767	589	9	9	9	15	15	15	38	38
murE	Cell division	cytosol	121.3	1488	496	5	5	5	10	10	10	47	47
murF	Cell division	cytosol	67.1	1359	453	7	7	7	12	12	12	34	34
mraY	Cell division	membrane	13.7	1083	361	5	5	5	9	9	9	16	16
murD	Cell division	cytosol	67.5	1317	439	3	3	3	12	12	12	36	36
ftsW	Cell division	membrane	2.7	1245	415	13	13	13	19	19	19	27	27
murG	Cell division	membrane	21.5	1068	356	5	5	5	11	11	11	34	34
murC	Cell division	cytosol	83.4	1476	492	2	2	2	12	12	12	29	29
ddlB	Cell wall synthesis	cytosol	33.1	921	307	7	7	7	15	15	15	24	24
ftsQ	Cell division	membrane	5.4	831	277	3	3	3	12	12	12	15	15
ftsA	Cell division	membrane	113.6	1263	421	10	10	10	9	9	9	23	23
ftsZ	Cell division	cytosol	633.6	1152	384	0	0	0	3	3	3	30	30
Total number of target codons:						83	83	83	157	157	157	374	374

b

Gene	Function	Localisation	Protein level ppm	ORF length	Peptide length	Number of target codons							
						r.s.1	r.s.2	r.s.3	r.s.4	r.s.5	r.s.6	r.s.7	r.s.8
rpmH	Protein translation	cytosol	4075.7	141	47	0	0	0	0	0	0	9	9
rpmD	Protein translation	cytosol	3046.3	180	60	0	0	0	2	2	2	20	20
rpmC	Protein translation	cytosol	5980.0	192	64	0	0	0	1	1	1	12	12
rpsR	Protein translation	cytosol	5806.5	228	76	0	0	0	0	0	0	2	2
rpmB	Protein translation	cytosol	5502.1	237	79	0	0	0	1	1	1	5	5
rpsP	Protein translation	cytosol	6611.2	249	83	3	3	3	2	2	2	16	16
rpsQ	Protein translation	cytosol	2179.3	255	85	1	1	1	0	0	0	3	3
rpmA	Protein translation	cytosol	4380.7	258	86	0	0	0	0	0	0	3	3
rpsS	Protein translation	cytosol	3094.6	279	93	0	0	0	0	0	0	3	3
rplW	Protein translation	cytosol	2091.7	303	101	0	0	0	1	1	1	7	7
rpsN	Protein translation	cytosol	4612.5	306	102	2	2	2	0	0	0	9	9
rplU	Protein translation	cytosol	1856.1	312	104	0	0	0	4	4	4	10	10
rpsJ	Protein translation	cytosol	3472.7	312	104	0	0	0	1	1	1	9	9
rplX	Protein translation	cytosol	4456.0	315	105	1	1	1	2	2	2	5	5
rplV	Protein translation	cytosol	7848.2	333	111	0	0	0	1	1	1	5	5
rplS	Protein translation	cytosol	3859.3	348	116	0	0	0	0	0	0	4	4
rplR	Protein translation	cytosol	6367.3	354	118	0	0	0	0	0	0	8	8
rplT	Protein translation	cytosol	3291.4	357	119	0	0	0	0	0	0	9	9
rpsM	Protein translation	cytosol	5733.1	357	119	1	1	1	0	0	0	10	10
rplL	Protein translation	cytosol	14543.5	366	122	0	0	0	0	0	0	8	8
rplN	Protein translation	cytosol	8866.6	372	124	1	1	1	2	2	2	3	3
rpsL	Protein translation	cytosol	5532.8	375	125	0	0	0	1	1	1	10	10
rplQ	Protein translation	cytosol	4272.8	384	128	0	0	0	1	1	1	5	5
rpsK	Protein translation	cytosol	2900.5	390	130	1	1	1	0	0	0	5	5
rpsH	Protein translation	cytosol	3828.3	393	131	0	0	0	0	0	0	1	1
rpsI	Protein translation	cytosol	3410.8	393	131	0	0	0	2	2	2	5	5
rplP	Protein translation	cytosol	3778.1	411	137	0	0	0	2	2	2	9	9
rplM	Protein translation	cytosol	4268.0	429	143	0	0	0	0	0	0	9	9
rplO	Protein translation	cytosol	5111.6	435	145	1	1	1	1	1	1	3	3
rplJ	Protein translation	cytosol	7731.6	498	166	1	1	1	2	2	2	3	3
rpsE	Protein translation	cytosol	7657.3	504	168	0	0	0	1	1	1	11	11
rplF	Protein translation	cytosol	5012.1	534	178	0	0	0	0	0	0	5	5
rpsG	Protein translation	cytosol	8660.2	540	180	0	0	0	2	2	2	11	11
rplE	Protein translation	cytosol	3489.1	540	180	0	0	0	2	2	2	5	5
rplD	Protein translation	cytosol	3469.9	606	202	0	0	0	1	1	1	5	5
rpsD	Protein translation	cytosol	5187.4	621	207	1	1	1	3	3	3	9	9
rplC	Protein translation	cytosol	4460.3	630	210	0	0	0	0	0	0	6	6
rpsC	Protein translation	cytosol	5755.0	702	234	0	0	0	0	0	0	2	2
rpsB	Protein translation	cytosol	4324.5	726	242	0	0	0	0	0	0	1	1
rplB	Protein translation	cytosol	5658.4	822	274	1	1	1	1	1	1	16	16
prfB	Protein translation	cytosol	570.9	1099	366	0	0	0	0	0	0	8	8
rpsA	Protein translation	cytosol	2649.1	1674	558	0	0	0	0	0	0	3	3
Total number of target codons:						14	14	14	36	36	36	292	292

Protein functions, localizations, expression levels (in parts per million), and lengths (both ORF length in bp and peptide length in amino acid count) of the genes in the essential cell division operon all simultaneously recoded in this work (a), and of individually recoded ribosomal and release factor 2 genes reported previously³⁷ (b). The numbers of codons targeted for removal according to different recoding schemes are also reported. The expression level data are from <http://www.pax-db.org>.

The formation of Charon's red poles from seasonally cold-trapped volatiles

W. M. Grundy¹, D. P. Cruikshank², G. R. Gladstone³, C. J. A. Howett⁴, T. R. Lauer⁵, J. R. Spencer⁴, M. E. Summers⁶, M. W. Buie⁴, A. M. Earle⁷, K. Ennico², J. Wm. Parker⁴, S. B. Porter⁴, K. N. Singer⁴, S. A. Stern⁴, A. J. Verbiscer⁸, R. A. Beyer^{2,9}, R. P. Binzel⁷, B. J. Buratti¹⁰, J. C. Cook⁴, C. M. Dalle Ore^{2,9}, C. B. Olkin⁴, A. H. Parker⁴, S. Protopapa¹¹, E. Quirico¹², K. D. Retherford³, S. J. Robbins⁴, B. Schmitt¹², J. A. Stansberry¹³, O. M. Umurhan², H. A. Weaver¹⁴, L. A. Young⁴, A. M. Zangari⁴, V. J. Bray¹⁵, A. F. Cheng¹⁴, W. B. McKinnon¹⁶, R. L. McNutt Jr¹⁴, J. M. Moore², F. Nimmo¹⁷, D. C. Reuter¹⁸, P. M. Schenk¹⁹ & the New Horizons Science Team*

A unique feature of Pluto's large satellite Charon is its dark red northern polar cap¹. Similar colours on Pluto's surface have been attributed² to tholin-like organic macromolecules produced by energetic radiation processing of hydrocarbons. The polar location on Charon implicates the temperature extremes that result from Charon's high obliquity and long seasons in the production of this material. The escape of Pluto's atmosphere provides a potential feedstock for a complex chemistry^{3,4}. Gas from Pluto that is transiently cold-trapped and processed at Charon's winter pole was proposed^{1,2} as an explanation for the dark coloration on the basis of an image of Charon's northern hemisphere, but not modelled quantitatively. Here we report images of the southern hemisphere illuminated by Pluto-shine and also images taken during the approach phase that show the northern polar cap over a range of longitudes. We model the surface thermal environment on Charon and the supply and temporary cold-trapping of material escaping from Pluto, as well as the photolytic processing of this material into more complex and less volatile molecules while cold-trapped. The model results are consistent with the proposed mechanism for producing the observed colour pattern on Charon.

The most prominent colour feature in New Horizons images of Charon is its reddish northern polar cap. Figure 1a combines the blue (400–550 nm), red (540–700 nm) and near-infrared (NIR; 780–975 nm) channels from the Multispectral Visible Imaging Camera (MVIC), part of New Horizons' Ralph remote sensing package^{5,6}. Figure 1b shows how reflectance in these three filters varies with latitude, averaged over the longitudes shown in Fig. 1a. Charon's surface gets darker and redder towards higher latitudes. Colour ratios of NIR/red and red/blue show similar latitude dependences to one another², suggesting a single-pigment material with increasing abundance towards the pole. Additional approach images are shown in Extended Data Fig. 1.

Longitudinal variability in the NIR/blue colour ratio is shown in Fig. 1c. At higher latitudes, colours are redder across all of the longitudes observed, although the trend is not perfectly uniform. Deviations may be related to local variations in topography or other parameters. The red coloration is interrupted by a few impact craters with diameters of several kilometres. Impacts that size occur rarely, probably much less frequently than once every million years⁷, so their existence implies that the red material must accumulate slowly (see Methods).

Infrared spectroscopy also supports slow accumulation. Spectra of Charon's pole are dominated by H₂O ice absorptions similar to spectra of lower-latitude regions². The data are consistent with up to 10% tholin

mixed with H₂O ice at the millimetre depths that are probed by the infrared observations, so tholin deposition must occur slowly enough that H₂O resupply or upward mixing of H₂O by impact gardening can compete.

Charon's south pole is currently in winter night. New Horizons observed the night-side with the Long Range Reconnaissance Imager (LORRI⁸) approximately 2.6 days after the closest approach, illuminated by Pluto-shine. The images reveal a decreasing brightness towards the pole (Fig. 2 and Extended Data Fig. 2) that cannot be attributed to the declining illumination alone, but requires a decreased albedo of the pole relative to equatorial latitudes (see Methods). The sunlit northern hemisphere in LORRI approach images shows a comparable albedo decline towards that pole (Fig. 2d).

Charon's surface temperature responds to solar forcing on diurnal (6.39 Earth days) and annual (248 Earth years) timescales. The high obliquity (currently 119°) causes polar latitudes to experience long periods of continuous darkness, during which they become extremely cold. Complicating the situation, the eccentricity of Pluto's heliocentric orbit, currently 0.253, results in a factor of 2.8 difference in the intensity of sunlight between perihelion and aphelion. To assess the thermal history of Charon's surface, we ran thermophysical models⁹ that track the diurnal and annual vertical heat flow into and out of Charon's surface at different latitudes, accounting for the present-day orbital parameters¹⁰ (see Methods). Figure 3a shows model thermal histories for four different northern hemisphere latitudes over the past few centuries for a nominal thermal inertia of 10 J m⁻² K⁻¹ s^{-1/2}, which is consistent with estimates of Charon's diurnal thermal inertia from Herschel Space Telescope observations¹¹. Models for bracketing thermal inertias (2.5 and 40 J m⁻² K⁻¹ s^{-1/2}) are shown in Extended Data Fig. 3, spanning the range reported from icy satellites¹² to Kuiper belt objects¹³. Charon's north pole experienced more than a century of continuous, extremely low temperatures from the late 1800s through to the spring equinox in 1989. Lower latitudes experienced briefer periods of continuous extreme cold, and also reached less extreme minimum temperatures.

The latest Charon year shown in Fig. 3a provides an incomplete picture of Charon's long-term thermal history because the pole precesses and the longitude of perihelion regresses^{10,14}, both on 3-million-year (3-Myr) timescales. For an idea of Charon's longer-term thermal history, we ran thermal models for previous Charon years, selecting one every 4 × 10⁵ yr over the past 3 Myr. Durations of continuous periods colder than a 25 K threshold temperature for cold-trapping

¹Lowell Observatory, Flagstaff, Arizona, USA. ²NASA Ames Research Center, Moffett Field, California, USA. ³Southwest Research Institute, San Antonio, Texas, USA. ⁴Southwest Research Institute, Boulder, Colorado, USA. ⁵National Optical Astronomy Observatory, Tucson, Arizona, USA. ⁶George Mason University, Fairfax, Virginia, USA. ⁷Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ⁸University of Virginia, Charlottesville, Virginia, USA. ⁹Carl Sagan Center, SETI Institute, Mountain View, California, USA. ¹⁰NASA Jet Propulsion Laboratory, La Cañada Flintridge, California, USA. ¹¹University of Maryland, College Park, Maryland, USA. ¹²Université Grenoble Alpes, CNRS, IPAG, Grenoble, France. ¹³Space Telescope Science Institute, Baltimore, Maryland, USA. ¹⁴Johns Hopkins University Applied Physics Laboratory, Columbia, Maryland, USA. ¹⁵Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona, USA. ¹⁶Washington University in St Louis, St Louis, Missouri, USA. ¹⁷University of California, Santa Cruz, California, USA. ¹⁸NASA Goddard Space Flight Center, Greenbelt, Maryland, USA. ¹⁹Lunar and Planetary Institute, Houston, Texas, USA.

*A list of participants and their affiliations appears at the end of the paper.

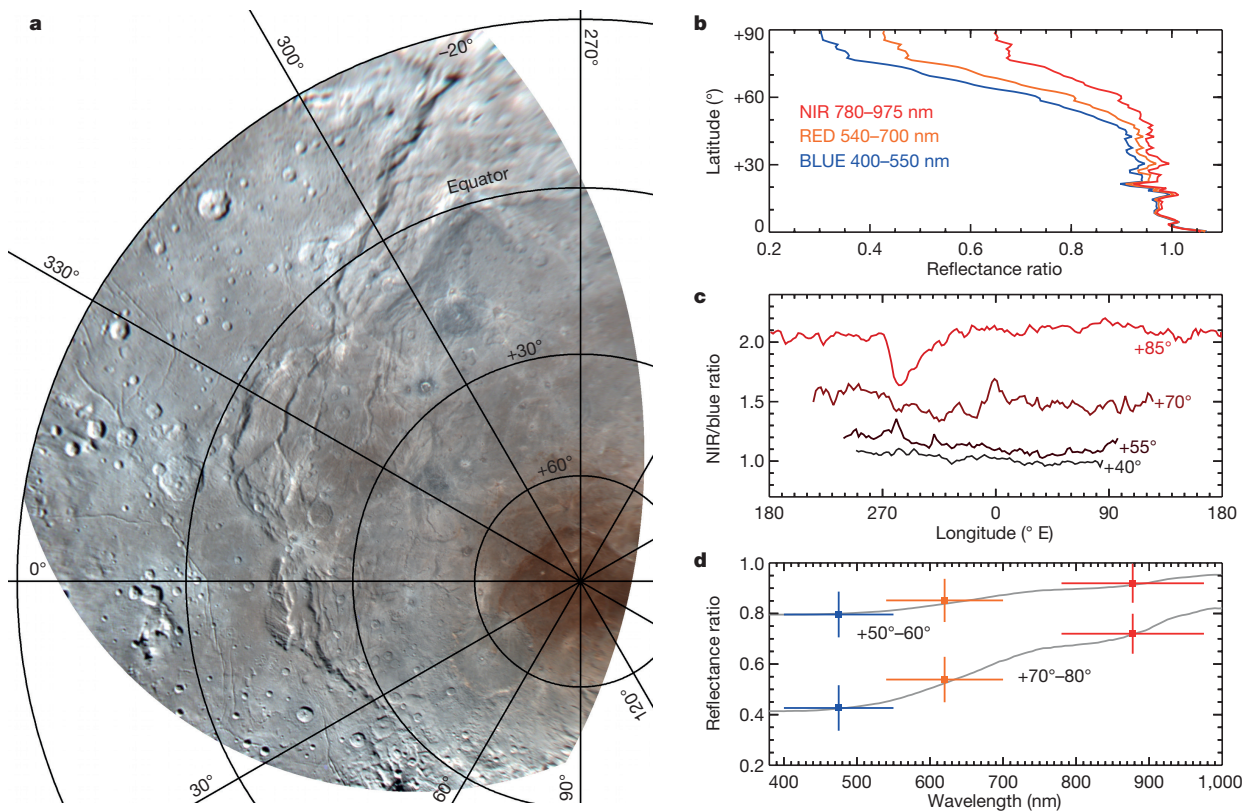


Figure 1 | Charon's red northern pole. **a**, Polar stereographic projection with Ralph's BLUE, RED and NIR filter images displayed in blue, green and red colour channels, respectively, relative to a Hapke photometric model (see Methods). **b**, Latitude dependence of the reflectance relative to the photometric model. **c**, Longitudinal dependence

of the NIR/BUE colour ratio. **d**, Wavelength dependence at two latitudes (coloured points) compared with spectral models of a laboratory tholin plus a neutral material (grey curves). The vertical bars indicate the standard deviation within each latitude bin; the horizontal bars indicate the filter widths.

(derived below) were averaged over the modelled epochs and also over the northern and southern hemispheres. These are shown as a function of latitude in Fig. 3b. Higher latitudes experience longer periods below the threshold temperature.

CH_4 and N_2 currently escape from Pluto at rates of 5×10^{25} and 1×10^{23} molecules s^{-1} , respectively, as estimated from New Horizons data⁴, with the Pluto exobase located at around 2.5 Pluto radii (that is, at an altitude of approximately 1,780 km)¹⁵. These are different from conditions assumed in a pre-encounter study³, which investigated the transfer of Pluto's escaping atmosphere to Charon including condensation on the winter pole. This model was based on an N_2 escape rate from Pluto of 2.3×10^{27} molecules s^{-1} , of which 5.7×10^{25} molecules s^{-1} (that is, around 2.5%) encountered Charon, leading to N_2 ice being deposited at a rate of $0.2 \mu\text{m}$ per decade on the winter pole. More recent simulations¹⁶ report comparable arrival fractions. The long-term temporal variability of the escape rate is not known.

Assuming that Charon intercepts 2.5% of the CH_4 escape flow determined by New Horizons data, that would correspond to a globally averaged arrival rate of 2.7×10^{11} molecules $\text{m}^{-2} \text{s}^{-1}$ at Charon. We estimate that most of the CH_4 remains at Charon long enough to find its way to the winter pole (see Methods). If it accumulates within 45° of the winter pole (a typical polar size, see Fig. 3b), it would produce a layer of CH_4 ice that is approximately $0.3 \mu\text{m}$ thick at each pole of Charon during the winter portion of the Pluto year. The accumulation would presumably vary with latitude, being thicker towards the pole.

Condensation on a surface depends on the temperature and vapour pressure¹⁷. The pressure at Charon's surface depends on the area of the polar cold trap. For a 45° radius pole, it can be estimated as 1×10^{-11} Pa. This pressure corresponds to an equilibrium vapour pressure of CH_4 ice at a temperature of 25 K (ref. 18). Where the surface temperature is colder, CH_4 will tend to freeze out as ice.

Methane escaping from Pluto's atmosphere is accompanied by other minor species. These include about 0.2% N_2 , various C_2 hydrocarbons (at a few tens of parts per million) and radicals such as CH_3 (at around 100 parts per million). N_2 is more volatile and thus requires lower temperatures to be cold-trapped, so it should freeze onto Charon's surface over a smaller range of latitudes and times of year than CH_4 does. Heavier hydrocarbons can condense anywhere on Charon, but would not produce enough of a deposit to be visible.

Our hypothesis requires energetic radiation to process the seasonally cold-trapped CH_4 . It is frozen on Charon's surface only during the polar winter night, so it must be processed rapidly, on the timescale of a century, and only by radiation impinging on the night side. It need not be fully converted into macromolecular solids such as tholins on such a short timescale, only into molecules that are sufficiently non-volatile to remain on the surface after the pole re-emerges into sunlight and warms back up. Charon's surface is subject to a variety of energetic radiation sources, including ultraviolet photons, solar wind charged particles, interstellar pickup ions and galactic cosmic rays^{19,20}. The most important night-side source of energetic radiation appears to be solar ultraviolet Lyman alpha photons ($\text{Ly}\alpha$, 10.2 eV) that have been scattered by the interplanetary medium, with a photon flux of $3.5 \times 10^{11} \text{ m}^{-2} \text{s}^{-1}$ on the night side. For 2.7×10^{11} molecules $\text{m}^{-2} \text{s}^{-1}$ arriving at Charon, concentration by cold-trapping in a 45° cold pole (approximately 1/7th of Charon's surface area) results in the accumulation of around 3 nm of ice per Earth year, of which we estimate 21% is photolysed (see Methods). There would probably also be some loss to sputtering²¹. N_2 ice is unaffected by $\text{Ly}\alpha$ ²² but some could be processed by cosmic rays.

When the winter pole re-emerges into sunlight, CH_4 and N_2 sublime away rapidly, but heavier, less volatile products remain behind. Assuming that the mass density of these photolytic products is double that of CH_4 ice, around 40 nm would accumulate

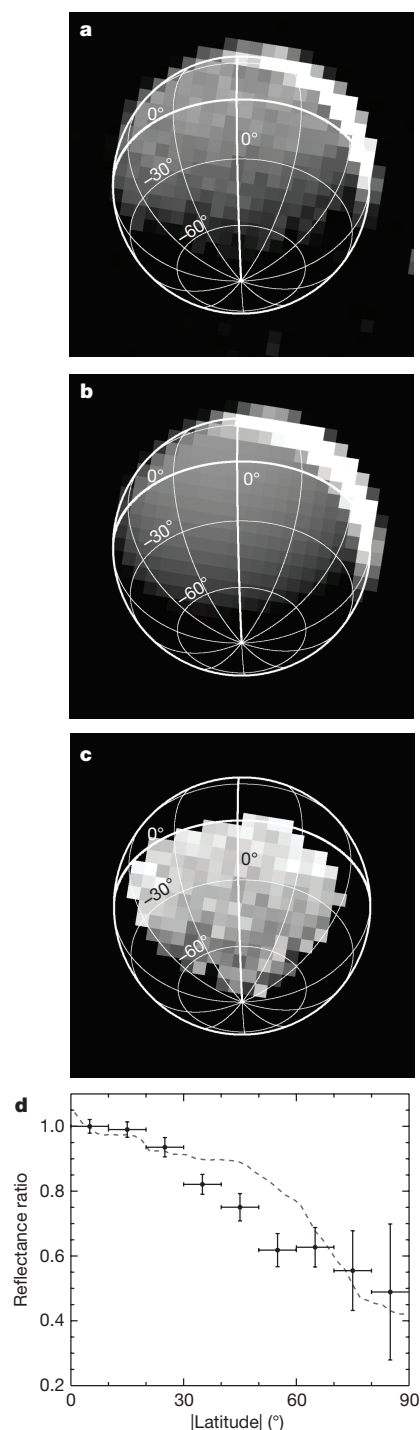


Figure 2 | Winter pole in Pluto-shine. **a**, Stack of 99 images showing a bright, sunlit crescent and fainter reflected light from Pluto. North, defined by the angular momentum vector, is up. **b**, Photometric model assuming a uniform albedo (see Methods). **c**, Observation/model ratio, showing that southern high latitudes are dark relative to equatorial latitudes. **d**, Dependence of the ratio on the absolute value of latitude as indicated by the *x* axis. The data points are for the Pluto-shine-illuminated southern hemisphere. The horizontal bars indicate the width of the latitude bin and the vertical bars show the standard deviation of the mean within each latitude bin. The dashed curve is for the sunlit northern hemisphere (see Methods).

per Pluto winter, or 0.16 mm per million Earth years. They will be exposed to other sources of energetic radiation including ultraviolet and extreme-ultraviolet photon radiation directly from the Sun, driving further photolytic chemistry²³ as well as sputtering erosion. At Pluto and Charon's mean heliocentric distance of 39 AU, the solar Ly α

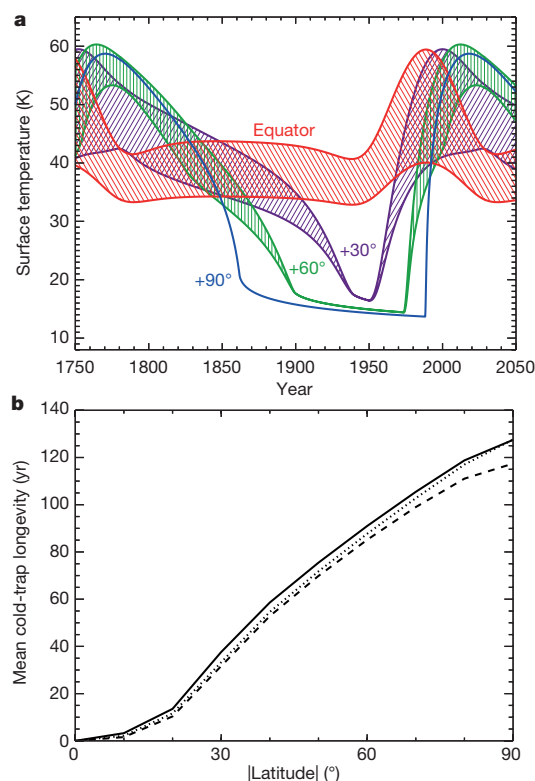


Figure 3 | Thermal environment. **a**, Model surface temperature history for the equator and three northern latitudes on Charon from 1750–2050 for a thermal inertia of $10 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$. Envelopes for each latitude indicate diurnal minimum and maximum temperatures. The dashed line is the 25 K threshold below which CH_4 is cold-trapped. **b**, Longest continuous duration below that temperature each Charon year, averaged over thermal models spanning the past 3 Myr. Dotted, solid, and dashed curves are for thermal inertias (in $\text{J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$) of 2.5, 10 and 40, respectively.

energy flux is $1.9 \times 10^{13} \text{ eV m}^{-2} \text{ s}^{-1}$, and that of the extreme-ultraviolet ($>12.4 \text{ eV}$) is $8.7 \times 10^{11} \text{ eV m}^{-2} \text{ s}^{-1}$. The charged particle flux from the ambient solar wind, coronal mass ejections and interstellar pickup ions is highly variable, and has energies ranging from a few electronvolts to megaelectronvolts. The production efficiency of tholin is not known for Charon's circumstances, but if it were 50% that would translate to around 30 cm of tholin produced at Charon's poles over 4 billion years.

Strongly coloured tholins have been made in the laboratory from a variety of CH_4 and N_2 ice mixtures using diverse radiation sources from charged particles to ultraviolet photons^{24,25}. For irradiation approximately $10 \times$ greater^{20,25}, experiments that initially produced a red-orange tholin generally go on to produce colourless and much darker materials, indicating carbonization (graphitization). The fact that Charon's pole is not completely blackened requires a balance between the production and further processing of tholins and processes such as in-falling H_2O ice dust, ejecta or micrometeorite impact gardening that would mix the tholins into the uppermost millimetres to few metres of Charon's H_2O regolith. Two distinct sources could contribute dust. Low-velocity dust from Pluto's small satellites is estimated to produce a 3 cm coating over the course of 4 billion years (ref. 26) with a composition that is probably dominated by H_2O ice. This is an order of magnitude slower than our estimated tholin production, although it could be augmented by ejecta from large impacts elsewhere on Charon. Higher-velocity Kuiper belt debris impact rates are poorly constrained, but models based on lunar rates^{27,28} imply gardening to centimetre depths on a timescale of 10^7 yr (see Methods), during which time we estimate about a millimetre of tholin accumulates. By diluting the accumulating tholin in local substrate material, this shallow gardening may explain the enduring brightness of a few relatively recent craters where accumulating tholin is mixed into more neutral-coloured H_2O -rich

ejecta, rather than being mixed into the already-darkened substrate, as elsewhere.

The distribution of dark, reddish material around Charon's northern pole is notable for its generally symmetric distribution across longitudes and its gradual increase with latitude, although there are local irregularities associated with craters, topographic features and perhaps subsurface variations in thermal properties. These characteristics, and the existence of an albedo feature around the southern pole with a similar latitude dependence, are consistent with our hypothesis that the combination of Pluto's escaping atmosphere and Charon's long, cold winters enables CH₄ to be seasonally cold-trapped at high latitudes, where some is photolytically processed into heavier molecules that are subsequently converted to reddish tholin-like materials. The symmetry argues against recent polar wander on Charon. Could the process occur elsewhere? Nix has a reddish spot²⁹, but it and the other small satellites orbit farther from Pluto and have much lower masses, making the process less efficient (see Methods).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 April; accepted 13 July 2016.

Published online 14 September 2016.

1. Stern, S. A. *et al.* The Pluto system: initial results from its exploration by New Horizons. *Science* **350**, aad1815 (2015).
2. Grundy, W. M. *et al.* Surface compositions across Pluto and Charon. *Science* **351**, aad9189 (2016).
3. Tucker, O. J., Johnson, R. E. & Young, L. A. Gas transfer in the Pluto–Charon system: a Charon atmosphere. *Icarus* **246**, 291–297 (2015).
4. Gladstone, G. R. *et al.* The atmosphere of Pluto as observed by New Horizons. *Science* **351**, aad8866 (2016).
5. Reuter, D. C. *et al.* Ralph: a visible/infrared imager for the New Horizons Pluto/Kuiper belt mission. *Space Sci. Rev.* **140**, 129–154 (2008).
6. Howett, C. J. *et al.* Inflight radiometric calibration of New Horizons' Multispectral Visible Imaging Camera (MVIC). Preprint at <http://arxiv.org/abs/1603.08940>.
7. Moore, J. M. *et al.* The geology of Pluto and Charon through the eyes of New Horizons. *Science* **351**, 1284–1293 (2016).
8. Cheng, A. F. *et al.* Long-Range Reconnaissance Imager on New Horizons. *Space Sci. Rev.* **140**, 189–215 (2008).
9. Spencer, J. R., Lebofsky, L. A. & Sykes, M. V. Systematic biases in radiometric diameter determinations. *Icarus* **78**, 337–354 (1989).
10. Earle, A. M. & Binzel, R. P. Pluto's insolation history: latitudinal variations and effects on atmospheric pressure. *Icarus* **250**, 405–412 (2015).
11. Lellouch, E. *et al.* The long-wavelength thermal emission of the Pluto–Charon system from Herschel observations: evidence for emissivity effects. *Astron. Astrophys.* **588**, A2 (2016).
12. Howett, C. J. A., Spencer, J. R., Pearl, J. & Segura, M. Thermal inertia and bolometric Bond albedo values for Mimas, Enceladus, Tethys, Dione, Rhea, and Iapetus as derived from Cassini/CIRS measurements. *Icarus* **206**, 573–593 (2010).
13. Lellouch, E. *et al.* "TNOs are Cool": A survey of the trans-Neptunian region IX. Thermal properties of Kuiper belt objects and Centaurs from combined Herschel and Spitzer observations. *Astron. Astrophys.* **557**, A60 (2013).
14. Dobrovolskis, A. R. & Peale, S. J. & Harris, A. W. in *Pluto and Charon* (eds Stern, S. A. & Tholen, D. J.) 159–190 (Univ. Arizona Press, 1997).
15. Bagenal, F. *et al.* Pluto's interaction: solar wind, energetic particles, dust. *Science* **351**, aad9045 (2016).
16. Hoey, W. A., Yeoh, S. K., Trafton, L. M., Goldstein, D. B. & Varghese, P. L. Rarefied gas dynamic simulation of transfer and escape in the Pluto–Charon system with the DSMC method. In *Proc. 47th Lunar and Planetary Science Conference* Contribution No. 1903 (Lunar and Planetary Institute, 2016).
17. Shorghofer, N. & Taylor, G. J. Subsurface migration of H₂O at lunar cold traps. *J. Geophys. Res.* **112**, E02010 (2007).
18. Fray, N. & Schmitt, B. Sublimation of ices of astrophysical interest: a bibliographic review. *Planet. Space Sci.* **57**, 2053–2080 (2009).
19. Johnson, R. E. *Energetic Charged-Particle Interactions with Atmospheres and Surfaces* (Springer, 1990).
20. Johnson, R. E. Effect of irradiation on the surface of Pluto. *Geophys. Res. Lett.* **16**, 1233–1236 (1989).
21. Mejía, C. *et al.* Compaction of porous ices rich in water by swift heavy ions. *Icarus* **250**, 222–229 (2015).
22. Cruz-Díaz, G. A., Muñoz Caro, G. M., Chen, Y. J. & Yih, T. S. Vacuum-ultraviolet spectroscopy of interstellar ice analogs II: absorption cross-sections of nonpolar ice molecules. *Astron. Astrophys.* **562**, A120 (2014).
23. Madey, T. E., Johnson, R. E. & Orlando, T. M. Far-out surface science: radiation-induced surfaces processes in the solar system. *Surf. Sci.* **500**, 838–858 (2002).
24. Materese, C. K., Cruikshank, D. P., Sandford, S. A., Imanaka, H. & Nuevo, M. Ice chemistry on outer solar system bodies: electron radiolysis of N₂, CH₄, and CO-containing ices. *Astrophys. J.* **812**, 150 (2015).

25. Thompson, W. R., Murray, B. G. J. P. T., Khare, B. N. & Sagan, C. Coloration and darkening of methane clathrate and other ices by charged particle irradiation: applications to the outer solar system. *J. Geophys. Res.* **92**, 14,933–14,947 (1987).
26. Porter, S. B. & Grundy, W. M. Ejecta transfer in the Pluto system. *Icarus* **246**, 360–368 (2015).
27. Gault, D. E., Hörz, F., Brownlee, D. E. & Hartung, J. B. Mixing of the lunar regolith. *Lunar Sci. Conf. Proc.* **3**, 2365–2386 (1974).
28. Spencer, J. R. & Denk, T. Formation of Iapetus' extreme albedo dichotomy by exogenically triggered thermal ice migration. *Science* **327**, 432–435 (2010).
29. Weaver, H. A. *et al.* The small satellites of Pluto as observed by New Horizons. *Science* **351**, aae0030 (2016).

Acknowledgements This work was supported by NASA's New Horizons Project. E.Q., B.S. and S.Phi. acknowledge the Centre National d'Etudes Spatiales (CNES) for its financial support through its 'Système Solaire' programme.

Author Contributions W.M.G. led the study and wrote the paper, with significant input from G.R.G. and M.E.S. on the escape of Pluto's atmosphere and its transport to Charon, D.P.C. on radiolytic production of tholins, C.J.A.H. and J.R.S. on thermal models, and T.R.L. on processing of Pluto-shine images. M.W.B., A.J.V., and B.J.B. developed photometric models. A.M.E. and R.P.B. computed insolation as functions of time and location. K.E. processed MVIC approach images. G.R.G., M.E.S., J.Wm.P., and K.D.R. assessed ultraviolet irradiation from various sources. K.N.S., S.B.P., and S.J.R. assessed effects of impacts from dust and larger projectiles. S.A.S. contributed diverse insights and led the overall mission. R.A.B., P.M.S., A.H.P., and A.M.Z. helped with geometric registration and image processing. C.M.D.O., J.C.C., S.Pro., C.B.O., E.Q., B.S., and S.Phi. interpreted infrared spectral data. J.A.S. and O.M.U. contributed insights on cold trapping of volatiles. R.L.M. contributed insights on the charged particle environment. L.A.Y. led development of the observation sequences that produced the data used here, and also checked the thermal models. H.A.W. and A.F.C. played key roles in LORRI instrument development and data processing, and D.C.R., C.B.O., A.H.P., and C.J.A.H. did the same for Ralph. W.B.M., J.M.M., K.N.S., F.N., P.M.S., and V.J.B. contributed insights on geophysics and geology of Charon.

Author Information All spacecraft data presented in this paper will be delivered to NASA's Planetary Data System (<https://pds.nasa.gov>) in a series of stages in 2016 and 2017 in accordance with the schedule established by NASA and the New Horizons project. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.G. (w.grundy@lowell.edu).

Reviewer Information *Nature* thanks L. Trafton and the other anonymous reviewer(s) for their contribution to the peer review of this work.

New Horizons Science Team

S. A. Stern¹, F. Bagenal², K. Ennico³, G. R. Gladstone⁴, W. M. Grundy⁵, W. B. McKinnon⁶, J. M. Moore³, C. B. Olkin¹, J. R. Spencer¹, H. A. Weaver⁷, L. A. Young¹, T. Andert⁸, O. Barnouin⁷, R. A. Beyer³, R. P. Binzel⁹, M. Bird¹⁰, V. J. Bray¹¹, M. Brozovi¹², M. W. Buie¹, B. J. Buratti¹², A. F. Cheng⁷, J. C. Cook¹, D. P. Cruikshank³, C. M. Dalle Ore^{13,3}, A. M. Earle⁹, H. A. Elliott⁴, T. K. Greathouse⁴, M. Hahn¹⁴, D. P. Hamilton¹⁵, M. E. Hill⁷, D. P. Hinson¹³, J. Hofgartner¹², M. Horányi², A. D. Howard¹⁶, C. J. A. Howett¹, D. E. Jennings¹⁷, J. A. Kammer¹, P. Kollmann⁷, T. R. Lauer¹⁸, P. Lavvas¹⁹, I. R. Linscott²⁰, C. M. Lisse², A. W. Lunsford¹⁷, D. J. McComas⁴, R. L. McNutt Jr.⁷, M. Mutchler²¹, F. Nimmo²², J. I. Nunez⁷, M. Paetzold¹⁴, A. H. Parker¹, J. Wm. Parker¹, S. Philippe²³, M. Piquette², S. B. Porter¹, S. Protopapa¹⁵, E. Quirico²³, H. J. Reitsema¹, D. C. Reuter¹⁷, S. J. Robbins¹, J. H. Roberts⁷, K. Runyon⁷, P. M. Schenk²⁴, E. Schindhelm¹, B. Schmitt²⁵, M. R. Showalter¹³, K. N. Singer¹, J. A. Stansberry²¹, A. J. Steffl¹, D. F. Strobel²⁵, T. Stryk²⁶, M. E. Summers²⁷, J. R. Szalay², H. B. Throop²⁸, C. C. C. Tsang¹, G. L. Tyler²⁰, O. M. Umurhan³, A. J. Verbisser¹⁶, M. H. Versteeg⁴, G. E. Weigle II⁴, O. L. White³, W. W. Woods²⁰, E. F. Young¹ & A. M. Zangari¹.

¹Southwest Research Institute, Boulder, Colorado, USA. ²University of Colorado, Boulder, Colorado, USA. ³NASA Ames Research Center, Moffett Field, California, USA. ⁴Southwest Research Institute, San Antonio, Texas, USA. ⁵Lowell Observatory, Flagstaff, Arizona, USA. ⁶Washington University in St. Louis, St. Louis, Missouri, USA. ⁷Johns Hopkins University Applied Physics Laboratory, Columbia, Maryland, USA. ⁸Universität der Bundeswehr München, Neubiberg 85577, Germany. ⁹Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ¹⁰University of Bonn, Bonn D-53113, Germany. ¹¹University of Arizona Lunar and Planetary Laboratory, Tucson Arizona, USA. ¹²NASA Jet Propulsion Laboratory, La Cañada Flintridge, California, USA. ¹³Carl Sagan Center at the SETI Institute, Mountain View, California, USA. ¹⁴Rheinisches Institut für Umweltforschung an der Universität zu Köln, Cologne 50931, Germany. ¹⁵University of Maryland, College Park, Maryland, USA. ¹⁶University of Virginia, Charlottesville, Virginia, USA. ¹⁷NASA Goddard Space Flight Center, Greenbelt, Maryland, USA. ¹⁸National Optical Astronomy Observatory, Tucson, Arizona, USA. ¹⁹Université de Reims Champagne-Ardenne, 51687 Reims, France. ²⁰Stanford University, Stanford, California, USA. ²¹Space Telescope Science Institute, Baltimore, Maryland, USA. ²²University of California, Santa Cruz, California, USA. ²³Université Grenoble Alpes, CNRS, IPAG, Grenoble, France. ²⁴Lunar and Planetary Institute, Houston, Texas, USA. ²⁵Johns Hopkins University, Baltimore, Maryland, USA. ²⁶Roane State Community College, Jamestown, Tennessee, USA. ²⁷George Mason University, Fairfax, Virginia, USA. ²⁸Planetary Science Institute, Tucson, Arizona, USA.

METHODS

Colour images. The colour image of Charon shown in Fig. 1a was obtained on 2015 July 14 at 10:42 UT. The spacecraft was 73,000 km from Charon, resulting in an MVIC image scale of 1.4 km per pixel. The Mission Elapsed Time (MET) unique label of this observation was 0299176432. Calibration details are described in ref. 6. The original image is shown in Extended Data Fig. 3c. Figure 1a shows this image reprojected on a polar stereographic projection and photometrically corrected by dividing each pixel by a Hapke model³⁰ computed for the same illumination and viewing geometry. That model had uniform parameters across the scene: single-scattering albedo $w = 0.9$, single-scattering phase function $P(g) = 0.8$, backscattering amplitude $B_0 = 0.6$, porosity parameter $h = 0.0044$ and macroscopic roughness $\theta = 20^\circ$ (parameters from ref. 31). Figure 1b shows the latitude dependence of the image divided by the Hapke model, averaged over the longitudes shown in Fig. 1a and normalized to unity at the equator. A similar latitude dependence in Charon's normal albedo has been reported from LORRI observations³². The longitude and wavelength dependence are shown in Fig. 1c, d, respectively. More approach images are shown in Extended Data Fig. 1, without reprojection or photometric correction.

Pluto-shine observations. New Horizons obtained 219 LORRI images of Charon's night side early on 2015 July 17 UT from a distance of approximately 3×10^6 km. The spacecraft's orientation is controlled by hydrazine thrusters, so full-resolution LORRI exposures must be short to minimize smear as pointing bounces around within a deadband. Longer exposures of 0.2 s were enabled by 4×4 pixel on-chip binning, resulting in a total integration of 44 s. The images were grouped into two sets, the first comprising 99 images with MET labels 0299398349 through 0299398716 acquired at a mean observation time of 00:26 UT. The second set had 120 images with MET labels 0299405549 through 0299405916 and a mean time of 02:26 UT. The two sets had image scales of 60 km per pixel and 62 km per pixel, respectively. As LORRI was pointed close to the Sun to observe the night side of Charon, the images were affected by scattered light patterns modified by small pointing variations among the images. The scattered light variability in each image set was modelled using principal component analysis on the complementary set, after masking Charon and bright stars from the images. The eigenimages in each set were used to subtract the scattered light contribution from the other set. The images were then co-registered at the sub-pixel level and combined into two stacks, the first of which is shown in Fig. 2a, the second in Extended Data Fig. 2a.

We modelled these observations by simulating Charon as a sphere approximated by 20,480 triangular facets. The bidirectional reflectance behaviour of each facet was represented with a Hapke model, using the same parameters as above. The Sun was treated as a point source and the distribution of light from Pluto was obtained by assuming it to be a Lambertian sphere with the appropriate size, location and illumination geometry. Being larger than Charon, Pluto casts some light onto Charon's pole, albeit obliquely. Comparisons between data and model are shown for the first of the two image stacks in Fig. 2 and for the other stack in Extended Data Fig. 2.

To compare Charon's southern pole with the northern one as seen with the same instrument, we also ran a Hapke model with the same parameters for a LORRI full-resolution sunlit approach image obtained 2015 July 14 at 02:44 UT, from a distance of 470,000 km, with image scale of 2.3 km per pixel. The MET label for this image was 299147776. It and the corresponding model are shown in Extended Data Fig. 2d and e and the latitude-dependent ratio is included as a dashed curve in Fig. 2d and Extended Data Fig. 2f, normalized to unity over the mean from 0° to 10° N. MVIC colour observations were not sufficiently sensitive to detect Pluto-shine on Charon's night side.

Thermal model. A standard one-dimensional finite element model⁹ was used to account for the vertical heat flow within each element of Charon's surface in response to diurnally and seasonally varying insolation¹⁰. Diurnal and annual timescales differ greatly, so we represented Charon's surface with a large number of layers (400) and broke its year into a large number of time steps (2×10^6 , about an hour per time step). We assumed a uniform bolometric bond albedo $A_B = 0.3$ and emissivity $\varepsilon = 0.9$, ignoring the lower albedos of Charon's poles. Diurnal thermal inertia Γ values were varied from 2.5 to $40 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$, spanning the range reported for Charon and Kuiper belt objects from Spitzer and Herschel observations^{11,13} to icy Saturnian satellites observed by Cassini¹². The density and heat capacity of H_2O ice are fixed, so low thermal inertias require inefficient conduction between ice grains. Conduction can rise with compaction at depth, leading to higher seasonal thermal inertias. Simulations with higher conductivities up to that of monolithic H_2O ice³³ at depths as shallow as few tens of centimetres below the surface can raise polar winter temperatures by a few kelvin without greatly affecting diurnal temperature variations. Sunlight was assumed to be the only energy source, providing 870 mW m^{-2} at Pluto's mean heliocentric distance. We did not include radiogenic heating. If Charon's rock fraction of around 60%

(ref. 34) has chondritic radionuclide abundances, heat from their decay would contribute an additional 1.5 mW m^{-2} , preventing the temperature from falling below about 13 K at present, and higher in the distant past. Figure 3a shows the model surface temperature history over the course of three centuries for Charon's northern hemisphere, assuming an intermediate diurnal thermal inertia value of $\Gamma = 10 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$, no radiogenic heat, and no high-conductivity subsurface layer. High and low Γ bounding cases are shown in Extended Data Fig. 3. Diurnal variations are much greater for small Γ , as is the range of seasonal surface temperature extremes. For all of these scenarios, Charon's high latitudes experience multi-decade episodes that are sufficiently cold to cold-trap CH_4 as ice.

As the spin pole of the system precesses and the longitude of the perihelion of its heliocentric orbit regresses^{10,14}, the insolation patterns evolve, modifying the surface temperature history over the course of a Pluto year. This variability is accommodated in Fig. 3b by averaging the longest continuous period below 25 K during a Pluto year over both hemispheres and over three million years of orbital history.

Loss mechanisms and cold-trapping. CH_4 molecules can be lost from Charon's surface environment by several mechanisms. First we consider thermal escape to space. For a Maxwell–Boltzmann distribution at 60 K, corresponding to the highest temperatures reached on Charon's summer hemisphere (see Fig. 3a), about 1% of CH_4 molecules exceed Charon's 590 m s^{-1} escape velocity. The majority, which do not escape hop on ballistic trajectories until they achieve escape velocity or encounter the cold pole and stick. A typical latitude boundary of the cold pole from Fig. 3b is 45° . A cold pole above that latitude occupies about 1/7th of Charon's surface. If each ballistic hop is assumed to arrive in a completely new random location, it would take an average of seven random hops to encounter the cold pole at a time when it is that size. With 1% lost to space per hop, 93% would make it to the pole, leading to an accumulation rate of $2 \times 10^{12} \text{ molecules m}^{-2} \text{ s}^{-1}$. For a CH_4 ice density³⁵ of 516 kg m^{-3} , the deposition rate is $9 \times 10^{-17} \text{ m s}^{-1}$ which adds up to about $0.3 \mu\text{m}$ at the pole over one Charon winter. Because the accumulation time is strongly latitude-dependent, the resulting ice distribution would be, too.

Another potential loss mechanism is ionization. If a molecule is ionized, it becomes coupled to the magnetic field in the solar wind and is thus swept from the system. The cross-section of a CH_4 gas molecule to $\text{Ly}\alpha$ radiation is estimated²² as $1.8 \times 10^{-21} \text{ m}^2$. The solar $\text{Ly}\alpha$ flux at Charon's mean heliocentric distance of 39 AU is $1.9 \times 10^{12} \text{ cm}^{-2} \text{ s}^{-1}$, so the probability of photoionization is about 0.1 per Earth year. At typical thermal speeds of $200\text{--}300 \text{ m s}^{-1}$, CH_4 molecules can traverse Charon's diameter in less than 2 h, so the probability of photoionization during seven ballistic hops is low. Likewise, little CH_4 would be photoionized between its escape from Pluto and arrival at Charon considering that Pluto's exobase temperature is around 70 K, so molecules escaping at the tail of the Maxwellian velocity distribution would reach Charon in just a few tens of hours.

Surface pressure. An estimate of the surface pressure at Charon can be obtained by assuming steady state, with arriving CH_4 molecules undergoing an average of seven ballistic hops before being lost to space or cold-trapped on a 45° radius cold pole, so each element of Charon's surface experiences seven times the globally averaged CH_4 arrival flux of $2.7 \times 10^7 \text{ molecules cm}^{-2} \text{ s}^{-1}$ at speeds that are consistent with the Boltzmann velocity distribution, resulting in a momentum flux or pressure of roughly $1 \times 10^{-11} \text{ Pa}$, which is three orders of magnitude smaller than the 3σ upper limit from New Horizons³⁶. When the cold pole is smaller, the pressure increases because more hops occur, and the reverse is true when the cold pole is larger, so this pressure is very approximate. However, it enables us to estimate the threshold temperature of 25 K for cold-trapping CH_4 via its vapour pressure¹⁸. Assuming the radius of a CH_4 molecule is $2 \times 10^{-10} \text{ m}$, this pressure corresponds to a mean free path that is >100 Charon radii, so collisions between CH_4 gas molecules can be ignored.

Photolysis of CH_4 ice. The interplanetary medium scatters $\text{Ly}\alpha$ photons from the Sun. The flux from this source observed during the New Horizons flyby⁴ was about 50% larger than predicted³⁷, or about 140 R averaged over the anti-Sun hemisphere, which illuminates the winter pole, equating to a flux of $3.5 \times 10^{11} \text{ Ly}\alpha$ photons $\text{m}^{-2} \text{ s}^{-1}$. This appears to be the largest source of night-side ultraviolet illumination at wavelengths between 10 nm and 133 nm that efficiently break bonds in the CH_4 molecule^{22,38}. Integrating over these wavelengths for the 1,000 stars with largest apparent ultraviolet fluxes using IUE spectra and Kurucz³⁹ models and dividing by two to account for the visible 2π steradians of sky contributes an additional $1.1 \times 10^{10} \text{ photons m}^{-2} \text{ s}^{-1}$. The Lambert absorption coefficient of CH_4 ice at the $\text{Ly}\alpha$ wavelength (1216 \AA)³⁸ is $19 \mu\text{m}^{-1}$ so a 53-nm-thick CH_4 ice coating has optical depth unity. The $\text{Ly}\alpha$ cross-section of individual CH_4 ice molecules has been reported²² as $1.4 \times 10^{-21} \text{ m}^2$, enabling an estimation of the probability of photolysis of an unshielded CH_4 ice molecule on Charon's winter pole as 1.5% per Earth year. For $2.7 \times 10^{11} \text{ molecules m}^{-2} \text{ s}^{-1}$ arriving at Charon, concentration by cold-trapping on approximately 1/7th of Charon's surface area results in the

accumulation of around 3 nm of CH₄ ice per Earth year, or a layer that is optically thick to Ly α in 18 yr. Only the skin reachable by ultraviolet photons can be photolysed. For steady state at this deposition rate, around 26% of the CH₄ would be processed before becoming buried beneath enough CH₄ ice to shield it from further ultraviolet photolysis. Over a century-long accumulation cycle at this rate, about 21% is processed.

Impacts and gardening. Only one of the bright-ray craters in Charon's polar region appears to be resolved. Its darker, inner region could mark the rim, or it could be ejecta given that many of Charon's other craters feature dark inner ejecta. If the dark/bright boundary is the rim, the diameter of the crater itself is around 5–6 km; otherwise, scaling from fully resolved craters, the diameter could be as small as approximately 1.5 km. In addition to uncertainty about crater size, the impactor flux in the outer Solar System for objects below around 100 km is not well known. Using the 'knee' model^{40,41} (consistent with the size distribution of larger craters on Pluto and Charon⁷), one crater 5–6 km or larger in diameter should typically form somewhere on Charon approximately every million years. The probability of occurrence in dark polar material is lower, in proportion to its smaller surface area. If the crater is smaller than 5–6 km, the occurrence frequency would be higher, but the knee model used predicts many more craters smaller than 10 km than are observed on Pluto or Charon⁷, suggesting that the timescale for the formation of the observed craters could be considerably longer than a million years.

No detailed modelling has so far been conducted for impact gardening of outer Solar System objects. Two studies of lunar impact gardening found similar results for 10⁷ yr timescales^{27,42}. A 50% probability that material has been excavated down to a depth of 1 cm over that interval was estimated²⁷, while ref. 42 found that all of the study points in a grid were excavated at least once down to a depth of around 1 cm. The two studies diverged for longer timescales, with ref. 42 deriving a disturbed depth of 10 cm in 10⁸ yr and the model of ref. 27 taking 10⁹ yr to achieve the same disturbed depth.

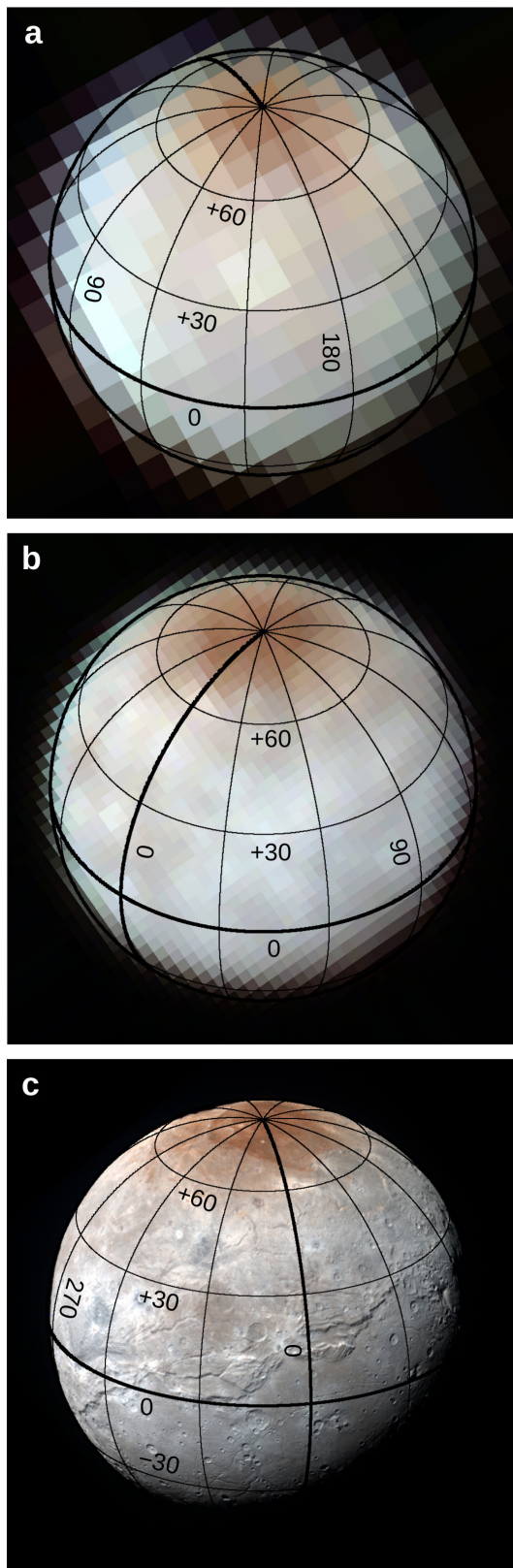
The rate of gardening in the lunar example is calculated for impactors of all sizes, from the largest objects to hit the Moon down to micrometeorites and dust-sized particles. Extrapolation to the outer Solar System requires knowledge of the impactor flux, which, as described above, is poorly constrained. The lunar flux used by ref. 27 is larger than the best estimate from refs 40 and 41 for Charon for impactors larger than around 0.1 mm by up to several orders of magnitude, but smaller than the extrapolated knee model for smaller impactors by similar amounts. The dust flux (particles below 1 cm in diameter) estimated from the New Horizons Student Dust Counter⁴³ is four to six orders of magnitude below the knee model. Extrapolating from that model to much smaller sizes or from the New Horizons dust measurement to much larger sizes is highly uncertain, illustrating the broad range of the potential impactor flux.

Other instances. A natural question to ask is whether our hypothesized mechanism for the production of Charon's dark, red poles should produce similar deposits elsewhere. Unlike the case for Charon, escape velocities on Pluto's smaller satellites are negligible compared with the thermal speeds. They intercept Pluto's escaping atmosphere only in proportion to their geometric cross-sections, with no enhancement effect from CH₄ molecules ballistically hopping around

their surfaces. We consider Nix, as it exhibits a prominent red spot²⁹, although similar calculations could be done for any of the small satellites. Nix's effective radius is 20 km and its orbital distance is 48,760 km, so it intercepts 1.7×10^{-7} of Pluto's escaping CH₄, or 8×10^{18} molecules s⁻¹. Nix's escape velocity is roughly 10–20 m s⁻¹, so only CH₄ molecules that happen to hit an extremely cold region on their first impact will stick. Assuming that Nix has a long-duration cold pole like Charon does (it might not, if its pole precesses rapidly), the polar orientation with respect to radially outflowing gas from Pluto would diminish the accumulation rate by another factor of four, leading to an accumulation rate of around 1×10^{-7} μ m yr, about 20,000 \times slower than the accumulation rate at Charon's cold pole. Such slow accumulation might not be competitive with impact erosion of Nix's surface.

Code availability. New Horizons MVIC and LORRI images were processed using the US Geological Survey's Integrated Software for Images and Spectrometers (ISIS), available at <https://isis.astrogeology.usgs.gov>. Thermal models were based on J.R. Spencer's thermprojs.pro, available at <https://www.boulder.swri.edu/~spencer/thermprojs>. Hapke reflectance models were based on M.W. Buie's bidr2.pro, available at <http://www.boulder.swri.edu/~buie/idl/pro/bidr2.html>.

30. Hapke, B. *Theory of Reflectance and Emittance Spectroscopy* (Cambridge Univ. Press, 1993).
31. Buie, M. W., Grundy, W. M., Young, E. F., Young, L. A. & Stern, S. A. Pluto and Charon with the Hubble Space Telescope I: monitoring global change and improved surface properties from light curves. *Astron. J.* **139**, 1117–1127 (2010).
32. Buratti, B. J. *et al.* Global albedos of Pluto and Charon from LORRI New Horizons observations. Preprint at <http://arxiv.org/abs/1604.06129> (2016).
33. Ross, R. G. & Kargel, J. S. in *Solar System Ices* (eds Schmitt, B. *et al.*) 33–62 (Kluwer Academic, 1998).
34. McKinnon, W. B. *et al.* The Pluto–Charon system revealed: geophysics, activity, and origins. In *Proc. 47th Lunar and Planetary Science Conference* Contribution No. 1995 (Lunar and Planetary Institute, 2016).
35. Bol'shutkin, D. N., Gasan, V. M. & Prokhvatilov, A. I. Temperature dependence of the crystalline lattice parameter of methane in the range of 11–70 °K. *J. Struct. Chem.* **12**, 670–672 (1972).
36. Stern, S. A. *et al.* New Horizons constraints on Charon's present day atmosphere. Preprint available at <http://arxiv.org/abs/1608.06955> (2016).
37. Gladstone, G. R., Pryor, W. R. & Stern, S. A. Ly α at Pluto. *Icarus* **246**, 279–284 (2015).
38. Baratta, G. A., Leto, G. & Palumbo, M. E. A comparison of ion irradiation and ultraviolet photolysis of CH₄ and CH₃OH. *Astron. Astrophys.* **384**, 343–349 (2002).
39. Kurucz, R. L. *SYNTHESIS Spectrum Synthesis Programs and Line Data* (Smithsonian Astrophysical Observatory, 1993).
40. Greenstreet, S., Gladman, B. & McKinnon, W. B. Impact and cratering rates onto Pluto. *Icarus* **258**, 267–288 (2015).
41. Greenstreet, S., Gladman, B. & McKinnon, W. B. Corrigendum to "Impact and cratering rates onto Pluto" [*Icarus* 258 (2015) 267–288]. *Icarus* **274**, 366–367 (2016).
42. Arnold, J. R. A Monte Carlo model for the gardening of the lunar regolith. *Moon* **13**, 159–172 (1975).
43. Poppe, A. R. Interplanetary dust influx to the Pluto–Charon system. *Icarus* **246**, 352–359 (2015).

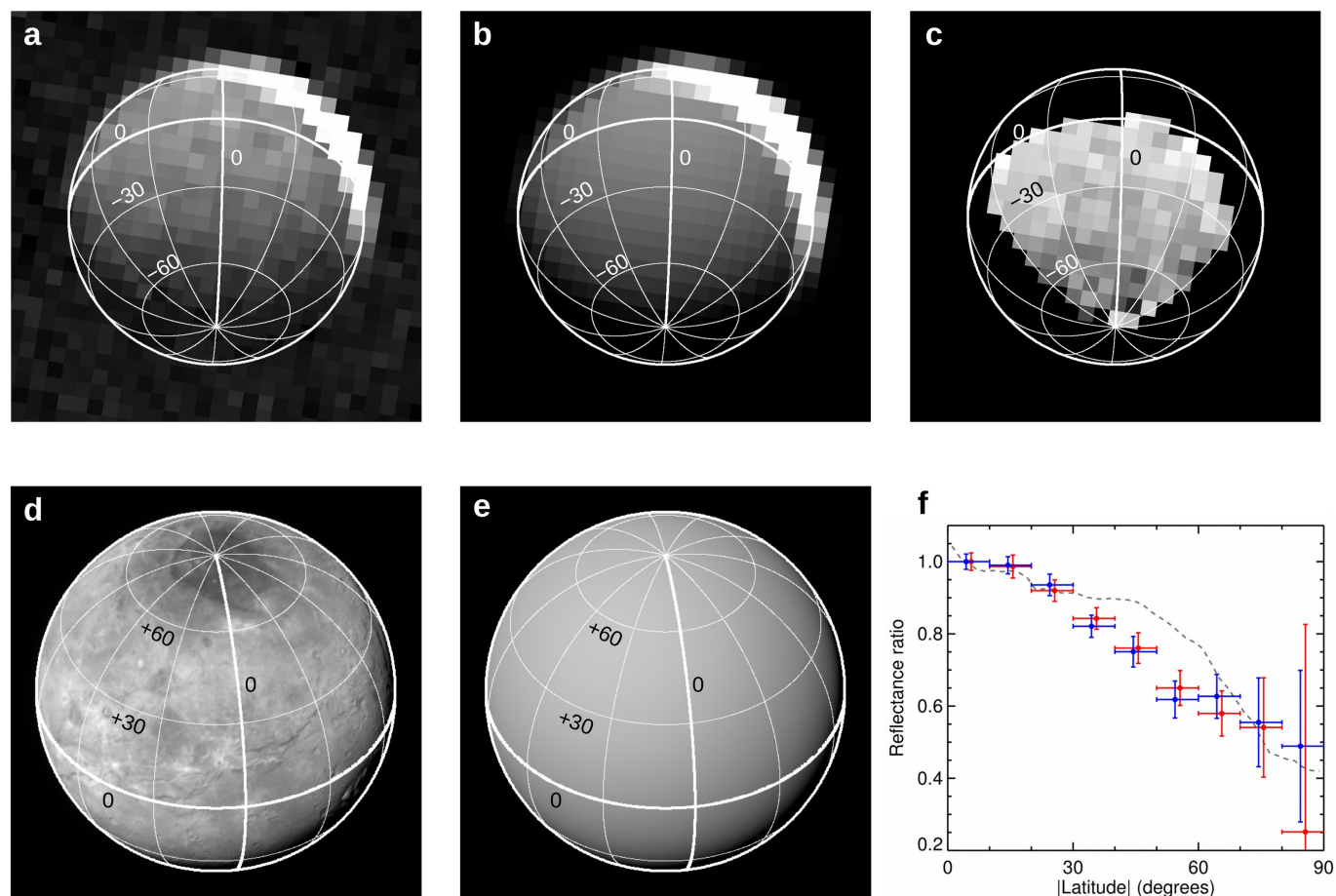


Extended Data Figure 1 | Three MVIC colour images obtained on approach showing Charon's northern pole as Charon rotates.

a, Observation obtained 2015 July 11 3:35 UT, with MET label 0298891582.

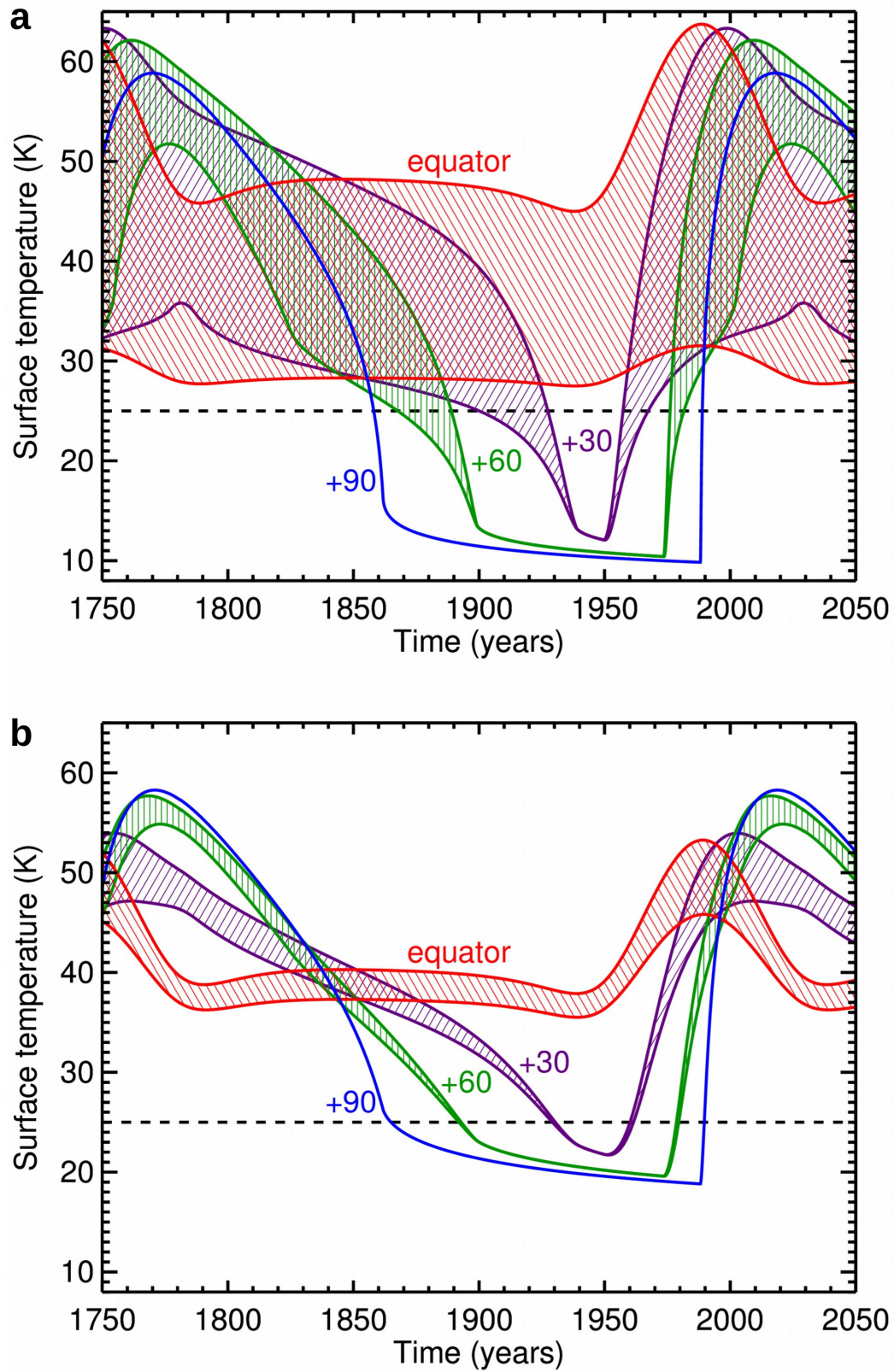
b, Observation obtained 2015 July 13 3:38 UT, with MET label 0299064592.

c, The same observation as Fig. 1a of the main text, obtained 2015 July 14 at 10:42 UT, with MET label 0299176432. Unlike in Fig. 1a, these images are not re-projected or divided by photometric models. North, as defined by the angular momentum vector, is up.



Extended Data Figure 2 | Additional panchromatic observations of Charon's poles. **a**, Second stack of 120 images of Charon's southern hemisphere illuminated by Pluto-shine obtained approximately 2 h after the stack in Fig. 2a. **b**, Corresponding photometric model. **c**, Observation/model ratio. **d**, Sunlit northern hemisphere. **e**, Corresponding photometric

model. **f**, Similar to Fig. 2d, with the first Pluto-shine stack indicated by blue points and the second stack indicated by red points (offset left and right for clarity). The horizontal bars indicate the widths of the latitude bins and the vertical bars indicate the standard deviation of the mean within each latitude bin.



Extended Data Figure 3 | Thermal models for cases of low and high thermal inertia. **a**, Thermal history for lower limit thermal inertia $\Gamma = 2.5 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$. **b**, Thermal history for upper limit thermal inertia $\Gamma = 40 \text{ J m}^{-2} \text{ K}^{-1} \text{ s}^{-1/2}$.

Calculation of the axion mass based on high-temperature lattice quantum chromodynamics

S. Borsanyi¹, Z. Fodor^{1,2,3}, J. Guenther¹, K. -H. Kampert¹, S. D. Katz^{3,4}, T. Kawanai², T. G. Kovacs⁵, S. W. Mages², A. Pasztor¹, F. Pittler^{3,4}, J. Redondo^{6,7}, A. Ringwald⁸ & K. K. Szabo^{1,2}

Unlike the electroweak sector of the standard model of particle physics, quantum chromodynamics (QCD) is surprisingly symmetric under time reversal. As there is no obvious reason for QCD being so symmetric, this phenomenon poses a theoretical problem, often referred to as the strong CP problem. The most attractive solution for this¹ requires the existence of a new particle, the axion^{2,3}—a promising dark-matter candidate. Here we determine the axion mass using lattice QCD, assuming that these particles are the dominant component of dark matter. The key quantities of the calculation are the equation of state of the Universe and the temperature dependence of the topological susceptibility of QCD, a quantity that is notoriously difficult to calculate^{4–8}, especially in the most relevant high-temperature region (up to several gigaelectronvolts). But by splitting the vacuum into different sectors and re-defining the fermionic determinants, its controlled calculation becomes feasible. Thus, our twofold prediction helps most cosmological calculations⁹ to describe the evolution of the early Universe by using the equation of state, and may be decisive for guiding experiments looking for dark-matter axions. In the next couple of years, it should be possible to confirm or rule out post-inflation axions experimentally, depending on whether the axion mass is found to be as predicted here. Alternatively, in a pre-inflation scenario, our calculation determines the universal axionic angle that corresponds to the initial condition of our Universe.

In this Letter, we use the lattice formulation of QCD¹⁰, that is, we discretize space-time on a four-dimensional lattice with N_t and N_s points in the temporal and spatial directions. The lattice spacing is denoted by a , the box size by $L = N_s a$, the temperature by $T = (aN_t)^{-1}$ and the volume of space-time by $V = N_s^3 N_t a^4$.

During the expansion of the early Universe, a QCD transition occurred that confined quarks and gluons into hadrons. Our most important qualitative knowledge about this transition is that it is an analytic crossover¹¹, thus no cosmological relics are expected. Outside the narrow temperature range of the transition we know that the Hubble rate and the relationship between temperature and the age of the early Universe can be described by a radiation-dominated equation of state (EoS). The calculation of the EoS is a challenging task, and the determination of the continuum limit at large temperatures is particularly difficult.

In our lattice QCD set-up, we used the staggered fermion discretization¹² with four steps of stout-smearing¹³. In our simulations, we included the two light quarks and the strange quark ('2 + 1 flavours') and when necessary we also added the charm quark ('2 + 1 + 1 flavours'). The quark masses are set to their physical values, but we use degenerate up and down quark masses and the small effect of isospin breaking is included analytically. The continuum limit is taken using three, four or five lattice spacings with temporal lattice extensions of $N_t = 6, 8, 10, 12$ and 16. In addition to dynamical staggered simulations,

we also used dynamical simulations with 2 + 1 flavours of overlap quarks¹⁴ down to physical masses. The inclusion of an odd number of flavours was a non-trivial task, but this set-up was required for the determination of the temperature dependence of the topological susceptibility of QCD, $\chi(T)$, at large temperatures in the several GeV region.

Charm quarks start to contribute to the EoS only above 300 MeV, so up to 250 MeV we used just 2 + 1 flavours of dynamical quarks. Connecting the 2 + 1 and the 2 + 1 + 1 flavour results at 250 MeV can be done smoothly. For large temperatures, the step-scaling method for the EoS of ref. 15 was applied. We determined the EoS with complete control over all sources of systematics all the way to the GeV scale.

We used two different methods to set the overall scale in order to determine the EoS. One of them took the pion decay constant, the other applied the w_0 scale¹⁶. Thirty-two different analyses (for example, the two different scale-setting procedures, different interpolations, keeping or omitting the coarsest lattice) entered our histogram method^{17,18} to estimate systematic errors. We also calculated the goodness of the fit and weights based on the Akaike information criterion, AICc¹⁸, and we looked at the unweighted or weighted results. This provided the systematic errors on our findings. In the low-temperature region, we compared our results with the prediction of the hadron resonance gas (HRG) approximation and found good agreement (within error bars). This HRG approach is used to parameterize the EoS for small temperatures. In addition, we used the hard thermal loop approach¹⁹ to extend the EoS to high temperatures.

In order to have a complete description of the thermal evolution of the early Universe, we supplement our QCD calculation for the EoS by including the rest of the standard model particles (leptons, bottom and top quarks, the photon, W , Z , Higgs bosons) and results on the electroweak transition. As a consequence, the final result on the EoS covers four orders of magnitude in temperature, from MeV to several hundred GeV. Figure 1 shows the EoS. The widths of the lines represent the uncertainties. Both the figure and the data can be used (similarly to figure 22.3 of ref. 20) to describe the Hubble rate and the relationship between temperature and the age of the Universe in a very broad temperature range.

We now turn to the determination of the other key quantity, $\chi(T)$. In general, the action of QCD should have a term proportional to the topological charge of the gluon field, Q . This term violates the combined charge-conjugation and parity (CP) symmetry. The surprising experimental observation is that the proportionality factor of this term θ is unnaturally small—this is known as the strong CP problem. A particularly attractive solution to this fundamental problem is the so-called Peccei–Quinn mechanism¹. An additional scalar U(1) symmetric field is introduced. The underlying Peccei–Quinn U(1) symmetry is spontaneously broken—which can happen pre-inflation or post-inflation—and an axion field A acts as a massless

¹Department of Physics, University of Wuppertal, D-42119 Wuppertal, Germany. ²Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52428 Jülich, Germany. ³Institute for Theoretical Physics, Eötvös University, H-1117 Budapest, Hungary. ⁴MTA-ELTE Lendület Lattice Gauge Theory Research Group, H-1117 Budapest, Hungary. ⁵Institute for Nuclear Research of the Hungarian Academy of Sciences, H-4026 Debrecen, Hungary. ⁶University of Zaragoza, E-50009 Zaragoza, Spain. ⁷Max Planck Institut für Physik, D-80803 Munich, Germany. ⁸Deutsches Elektronen-Synchrotron DESY, D-22607 Hamburg, Germany.

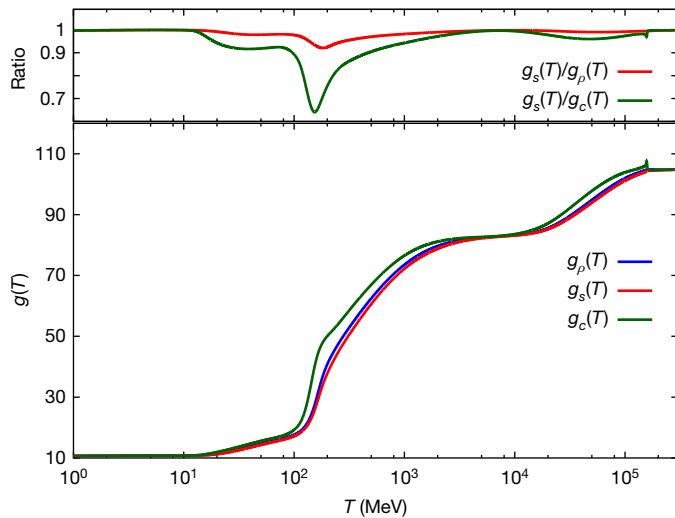


Figure 1 | The effective degrees of freedom of the energy density, the entropy density and the heat capacity in the early Universe, and their ratios. Shown as a function of temperature T in the lower panel are the degrees of freedom as follows: g_ρ for the energy density ($\rho = g_\rho \frac{\pi^2}{30} T^4$), g_s for

the entropy density ($s = g_s \frac{2\pi^2}{45} T^3$), and g_c for the heat capacity ($c = g_c \frac{2\pi^2}{15} T^3$).

Neglecting the cosmological constant, the time dependence of the temperature in the early Universe is given by these factors as:

$\frac{dT}{dt} = -\frac{2\pi^{3/2}}{3\sqrt{5}} \frac{T^3}{M_{\text{Pl}}} \frac{\sqrt{g_\rho g_s}}{g_c}$, where M_{Pl} is the Planck mass. The line width is chosen to be the same as our error bars (s.e.m.) at the vicinity of the QCD transition ($T \approx 150$ MeV), where we have the largest uncertainties. At temperatures $T < 1$ MeV the equilibrium EoS becomes irrelevant for cosmology, because of neutrino decoupling. The EoS comes from our calculation up to $T = 100$ GeV. At higher temperatures the electroweak transition becomes relevant and we use the results of ref. 22. Note that for temperatures around the QCD transition, non-perturbative QCD effects push the EoS away from the ideal gas limit, an approximation which is often used in cosmology; for example, g_s/g_c is reduced from this limit by about 35% (see upper panel). Also note that g_s/g_c has four local minima: near the muon threshold, the QCD transition, the W , Z -boson thresholds and the electroweak transition. For parameterizations for the QCD regime or for the whole temperature range see Supplementary Information.

Goldstone boson of the broken symmetry^{2,3}. The symmetry-breaking scale f_A is a free parameter. Owing to the chiral anomaly, the axion is coupled to the topological charge density, so the original potential of the axion field with its $U(1)$ symmetry breaking gets tilted and has its minimum where $(\theta + A/f_A) = 0$. This sets the proportionality factor of Q in the QCD action to zero and solves the strong CP problem. Furthermore, the axion acquires a mass m_A , which is given by $m_A^2 = \chi/f_A^2$, and $\chi = \langle Q^2 \rangle/V$ is the susceptibility of the topological charge normalized by the space-time volume. We determined its value at $T=0$, which was $\chi(T=0) = 0.0245(24)(12) \text{ fm}^{-4}$ in the isospin symmetric case; the first error in parentheses is statistical, the second is systematic. Isospin breaking results in a small, 12% correction, thus the physical value is $\chi(T=0) = 0.0216(21)(11) \text{ fm}^{-4} = [75.6(1.8)(0.9) \text{ MeV}]^4$.

On the lattice, χ can be conveniently calculated using a Q defined along the Wilson flow²¹. An earlier study⁵ looked at $\chi(T)$ in the quenched approximation. A result was provided within the quenched framework, and reached a temperature about one-half to one-third of the necessary temperatures for axion cosmology (a similar study with somewhat less control over the systematics is in ref. 4). To obtain a complete result, dynamical quarks with physical masses should be used. Dynamical configuration production is, however, about three orders of magnitude more expensive computationally, and the $\chi(T)$ values are several orders of magnitude smaller than in the quenched case. Owing to cut-off effects, the continuum limit is far more difficult to carry out in dynamical QCD than in the pure gauge theory⁵. All in

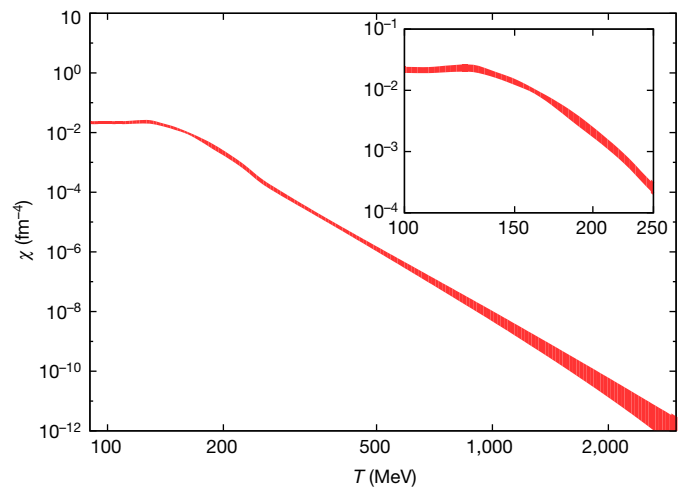


Figure 2 | Continuum limit of $\chi(T)$. Main panel, plot of the temperature dependence of the topological susceptibility of QCD, $\chi(T)$. The width of the line represents the combined statistical and systematic errors (s.e.m.). The dilute instanton gas approximation (DIGA) predicts a power behaviour of T^{-b} with $b = 8.16$, and the lattice result shown here is close to this value. Inset, magnified view of the behaviour around the QCD transition.

all, we estimate that the brute-force approach to providing a complete result on $\chi(T)$ in the relevant temperature region would be at least ten orders of magnitude more expensive than the result of ref. 5.

The huge computational demand and the physics issue behind the determination of $\chi(T)$ have two main sources: (a) in high-temperature lattice QCD, the most widely used actions are based on staggered quarks, and when dealing with topological observables staggered quarks have very large cut-off effects; and (b) the tiny topological susceptibility needs extremely long simulation threads to observe enough changes of the topological sectors.

We solve both problems and determine the continuum result for $\chi(T)$ for the entire temperature range of interest. We call our proposed solution of problem (a) the ‘eigenvalue reweighting’ technique. The method is based on substituting the topology related eigenvalues of the staggered quark operator with the eigenvalues of the quark operator in the continuum. To solve problem (b), we propose to measure the logarithmic differential of the susceptibility instead of the susceptibility itself, which is related to quantities that are to be measured in fixed topological sectors. The final result is obtained with an integral, so we call our method the ‘fixed sector integral’ technique. Both techniques are explained in detail in the Methods.

The CPU costs of the conventional technique scale as T^8 , whereas the new ‘fixed sector integral’ method scales as T^0 , so the reduction in CPU time is tremendous. This efficient technique is used to obtain the final result for $\chi(T)$. Since we work with continuum extrapolated quantities both for the ratios in the starting-point and for their changes, we could in principle use any action in the procedure: here we will use overlap and/or staggered actions.

By combining these methods, $\chi(T)$ can be determined (see Fig. 2). The cut-off effects of staggered fermions (several thousand per cent) are removed, leaving a very mild (of the order of 10%) continuum extrapolation to be performed. In addition, the direct determination of $\chi(T)$ all the way up to 3 GeV means that we do not have to rely on the dilute instanton gas approximation (DIGA). Note that *a posteriori* the exponent predicted by DIGA turned out to be compatible with our finding, but its prefactor is off by an order of magnitude, similar to the quenched case. Though some of our simulations (see Supplementary Fig. 18) are already carried out with chiral (overlap) fermions, where large cut-off effects are *a priori* absent, it is an important task for the future to crosscheck these results with a calculation using chiral fermions only.

As a possible application for these two cosmologically relevant lattice QCD results, we show how to calculate the amount of axionic

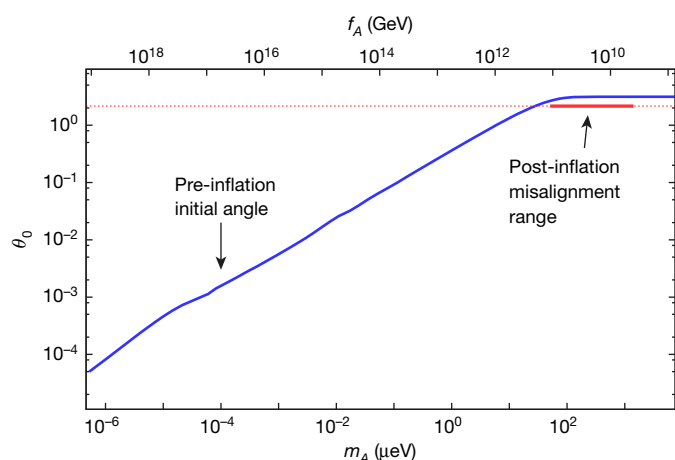


Figure 3 | The relation between the axion's mass m_A and the initial angle θ_0 in the pre-inflation scenario and the axion's mass range in the post-inflation scenario. For the pre-inflation scenario, our result is shown by the blue line; the error (s.e.m.) is smaller than the line width. The post-inflation scenario corresponds to $\theta_0 = 2.155$ with a strict lower bound on the axion's mass of $m_A = 28(2) \mu\text{eV}$. The thick red line shows our result for the axion's mass for the post-inflation case: for example, $m_A = 50(4) \mu\text{eV}$ if one assumes that axions from the misalignment mechanism contribute 50% to dark matter. Our final estimate is $50 \mu\text{eV} < m_A < 1,500 \mu\text{eV}$ (the upper bound assumes that only 1% is the contribution from the misalignment mechanism, the rest comes from other sources—for example, topological defects). An experimental set-up to detect post-inflationary axions is given in Supplementary Information. The slight bend around $m_A \approx 10^{-5} \mu\text{eV}$ corresponds to an oscillation temperature at the QCD transition^{23,24}.

dark matter and how it can be used to determine the axion's mass. $\chi(T)$ is a rapidly decreasing function of the temperature. Thus, at high temperature m_A (which is proportional to $\chi(T)^{1/2}$) is small: in fact, it is much smaller than the Hubble expansion rate of the Universe at that time or temperature ($H(T)$). The axion does not yet 'feel' the tilt in the Peccei–Quinn 'Mexican hat'-type potential, and it is effectively massless and frozen by the Hubble friction (a linear, friction-like term proportional to the Hubble constant in the equation of motion, see the second term of equation S29 in Supplementary Information). As the Universe expands, the temperature decreases, $\chi(T)$ increases and the axion mass also increases; meanwhile, the Hubble expansion rate (given by our EoS) decreases. As the temperature decreases to T_{osc} , the axion mass is of the same order as the Hubble constant (the oscillation temperature T_{osc} is defined by $3H(T_{\text{osc}}) = m_A(T_{\text{osc}})$). Around this time the axion field rolls down the potential, and starts to oscillate around the tilted minimum; the axion number density increases to a non-zero value, thus axions as dark matter are produced. The details of this production mechanism, usually called misalignment, are quite well known (see, for example, ref. 9).

In a post-inflationary scenario, the initial value of the angular degree of freedom (θ) of the axion field takes all values between $-\pi$ and π , whereas in the pre-inflationary scenario only one angle, θ_0 , contributes (all other values are inflated away). One should also mention that during the U(1) symmetry-breaking, topological strings appear, which decay and also produce dark-matter axions. In the pre-inflationary scenario they are inflated away. However, in the post-inflationary framework their role is more important. This sort of axion production mechanism is less well-understood, and in our final results it is necessary to make some assumptions about the amount of axions produced by topological strings (see Fig. 3 legend).

The direct consequence of our results on χ and the EoS is the mass of the dark-matter candidate, the axion. For the pre-inflationary Peccei–Quinn symmetry-breaking scenario, the initial value of the axion field of our Universe (θ_0) determines the axion mass (and vice versa). In Fig. 3 we show this relationship between the axion's mass m_A (and the symmetry-breaking scale f_A) and the initial angle θ_0 . For

the post-inflationary Peccei–Quinn scenario, the horizontal thick red interval shows the possible range on m_A .

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 June; accepted 12 September 2016.

- Peccei, R. D. & Quinn, H. R. CP conservation in the presence of instantons. *Phys. Rev. Lett.* **38**, 1440–1443 (1977).
- Weinberg, S. A new light boson? *Phys. Rev. Lett.* **40**, 223–226 (1978).
- Wilczek, F. Problem of strong P and T invariance in the presence of instantons. *Phys. Rev. Lett.* **40**, 279–282 (1978).
- Berkowitz, E., Buchoff, M. I. & Rinaldi, E. Lattice QCD input for axion cosmology. *Phys. Rev. D* **92**, 034507 (2015).
- Borsanyi, S. *et al.* Axion cosmology, lattice QCD and the dilute instanton gas. *Phys. Lett. B* **752**, 175–181 (2016).
- Trunin, A., Burger, F., Ilgenfritz, E.-M., Lombardo, M. P. & Müller-Preussker, M. Topological susceptibility from $N_f = 2 + 1 + 1$ lattice QCD at nonzero temperature. *J. Phys. Conf. Ser.* **668**, 012123 (2016).
- Bonati, C. *et al.* Axion phenomenology and θ -dependence from $N_f = 2 + 1$ lattice QCD. *J. High Energy Phys.* **2016**, 155 (2016).
- Petreczky, P., Schadler, H.-P. & Sharma, S. The topological susceptibility in finite temperature QCD and axion cosmology. *Phys. Lett. B* <http://dx.doi.org/10.1016/j.physletb.2016.09.063> (2016).
- Wentz, O. & Shellard, E. P. S. Axion cosmology revisited. *Phys. Rev. D* **82**, 123508 (2010).
- Wilson, K. G. Confinement of quarks. *Phys. Rev. D* **10**, 2445–2459 (1974).
- Aoki, Y., Endrodi, G., Fodor, Z., Katz, S. D. & Szabó, K. K. The order of the quantum chromodynamics transition predicted by the standard model of particle physics. *Nature* **443**, 675–678 (2006).
- Kogut, J. B. & Susskind, L. Hamiltonian formulation of Wilson's lattice gauge theories. *Phys. Rev. D* **11**, 395–408 (1975).
- Morningstar, C. & Peardon, M. J. Analytic smearing of SU(3) link variables in lattice QCD. *Phys. Rev. D* **69**, 054501 (2004).
- Neuberger, H. Exactly massless quarks on the lattice. *Phys. Lett. B* **417**, 141–144 (1998).
- Borsanyi, Sz., Endrodi, G., Fodor, Z., Katz, S. D. & Szabo, K. K. Precision SU(3) lattice thermodynamics for a large temperature range. *J. High Energy Phys.* **2012**, 56 (2012).
- Borsanyi, S. *et al.* High-precision scale setting in lattice QCD. *J. High Energy Phys.* **2012**, 10 (2012).
- Dürr, S. *et al.* Ab initio determination of light hadron masses. *Science* **322**, 1224–1227 (2008).
- Borsanyi, S. *et al.* Ab initio calculation of the neutron-proton mass difference. *Science* **347**, 1452–1455 (2015).
- Andersen, J. O., Leganger, L. E., Strickland, M. & Su, N. NNLO hard-thermal-loop thermodynamics for QCD. *Phys. Lett. B* **696**, 468–472 (2011).
- Olive, K. A. *et al.* Review of particle physics. *Chin. Phys. C* **38**, 090001 (2014).
- Lüscher, M. Properties and uses of the Wilson flow in lattice QCD. *J. High Energy Phys.* **2010**, 71 (2010); Erratum. *J. High Energy Phys.* **2014**, 92 (2014).
- Laine, M. & Meyer, M. Standard Model thermodynamics across the electroweak crossover. *J. Cosmol. Astropart. Phys.* **2015**, 035 (2015).
- Aoki, Y. *et al.* The QCD transition temperature: results with physical masses in the continuum limit II. *J. High Energy Phys.* **2009**, 088 (2009).
- Borsanyi, S. *et al.* Is there still any Tc mystery in lattice QCD? Results with physical masses in the continuum limit III. *J. High Energy Phys.* **2010**, 073 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Dierigl, M. Giordano, S. Krieg, D. Nogradi and B. Both for discussions. This project was funded by the DFG (grant SFB/TR55) and by OTKA (grant OTKA-K-113034). T.G.K. is supported by the Hungarian Academy of Sciences under 'Lendület' grant no. LP 2011-011. The work of J.R. is supported by the Ramon y Cajal Fellowship 2012-10597 and by FPA2015-65745-P (MINECO/FEDER). The computations were performed on JUQUEEN at Forschungszentrum Jülich, on SuperMUC at Leibniz Supercomputing Centre in München, on Hazel Hen at the High Performance Computing Center in Stuttgart, on QPACE in Wuppertal and on GPU clusters in Wuppertal, Budapest and Debrecen.

Author Contributions S.B. and S.W.M. developed the fixed sector integral, T.G.K. and K.K.S. the eigenvalue reweighting, and F.P. the odd flavour overlap techniques. S.B., J.G., S.D.K., T.G.K., T.K., S.W.M., A.P., F.P. and K.K.S. wrote the necessary codes, carried out the runs and determined the EoS and $\chi(T)$. A.P., J.R. and A.R. calculated the DIGA prediction. K.-H.K., J.R. and A.R. worked out the experimental set-up. Z.F. wrote the main paper and coordinated the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.F. (fodor@bodri.elte.hu).

Reviewer Information Nature thanks M. Paola and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Eigenvalue reweighting technique. Here we show how cut-off effects in χ arise with staggered quarks and propose a new method to efficiently suppress them.

The cut-off effects are strongly related to the zero modes. To understand their importance, we first note that in the quark determinant every zero mode for each dynamical flavour contributes a factor m_f the corresponding quark mass. In this way gauge configurations with zero modes are strongly suppressed in the path integral, especially if the quark masses are small. Owing to the index theorem, this also implies that light dynamical quarks strongly suppress higher topological sectors and thus χ itself.

On the lattice, however, there can be strong cut-off effects in this suppression. This is because the suppression factor is not m_f but $m_f + i\lambda_0$, where λ_0 is an eigenvalue corresponding to the would-be zero mode of the staggered Dirac operator, D_{st} . The lack of exact zero modes can thus introduce strong cut-off effects and slow convergence to the continuum limit. Indeed, as long as the typical would-be zero eigenvalues are comparable to or larger than the lattice bare quark mass m_f , higher sectors are much less suppressed on the lattice than in the continuum.

To improve the situation, even at finite lattice spacing we can identify the would-be zero modes and restore their continuum weight in the path integral. In case of rooted staggered quarks this amounts to a reweighting of each configuration (U) with a weight factor

$$w[U] = \prod_f \prod_{n=1}^{2|Q[U]|} \prod_{\sigma=\pm} \left(\frac{m_f}{\sigma i \lambda_n[U] + m_f} \right)^{n_f/4} \quad (1)$$

where the second product runs over the would-be zero eigenvalues (λ_n) of the staggered Dirac operator with positive imaginary part. For n_f fermion flavours the third product takes into account the $i\lambda \rightarrow -i\lambda$ symmetry of the eigenvalue spectrum. The $n_f/4$ factor takes rooting into account, the factor 2 next to $|Q|$ together with the \pm symmetry make up for the fact that in the continuum limit the staggered zero modes become fourfold degenerate²⁵.

We now turn to the most important part of the reweighting: the definition of the would-be zero modes. Since we are interested in χ , we identify the number of these modes with the magnitude of the topological charge $2|Q|$ as obtained from the gauge field after using the Wilson flow (see Supplementary Information). We investigated two specific choices for the would-be zero modes. In the first approach, we took the $2|Q|$ eigenmodes that have the largest magnitude of chirality among the eigenmodes with the appropriate sign of chirality, positive if $Q < 0$ and negative if $Q > 0$. In the second approach we took the $2|Q|$ eigenmodes with smallest magnitude. These two approaches are equivalent in the continuum limit, where zero modes are exactly at zero and their chirality is unity. In practical simulations they give very similar results; we use the second approach in our analysis.

Since in the continuum limit the would-be zero eigenvalues get closer to zero, the reweighting factors tend to unity and in the continuum limit we recover the original Dirac operator. This way, however, even at finite lattice spacings the proper suppression of higher sectors is restored and cut-off effects are strongly reduced resulting in much faster convergence to the continuum limit. For completeness, we note that the above modification corresponds to a non-local modification of the path integral. (In this respect it stands on a footing similar to another method, which also modifies the quark determinant and which we also use in our staggered simulations: determinant rooting. As of today there is ample theoretical and numerical evidence for the correctness of the staggered rooting. See ref. 26 and its follow-ups.) In the following we provide several pieces of numerical evidence for the correctness of the approach.

In Extended Data Fig. 1 we plot the distribution of the eigenvalues corresponding to the would-be zero modes at a temperature of $T = 240$ MeV for different lattice spacings. The distributions get narrower and their centre moves towards zero as the lattice spacing is decreased. In Extended Data Fig. 2 we show the expectation value of the reweighting factors in the first few sectors. In the continuum limit $\langle w \rangle_Q = 1$ should be fulfilled in each sector. The results nicely converge to 1.

In most of our runs, especially at large temperatures and small quark masses, the weights were much smaller than 1. As a result there are orders of magnitude differences between χ with and without reweighting. It is therefore important to illustrate how the standard approach breaks down if the lattice spacing is large and how the correct result is recovered for very small lattice spacings. In the following we show two examples, Extended Data Figs 3 and 4, where the standard method produces cut-off effects so large that a reliable continuum extrapolation is not possible. In contrast, the lattice spacing dependence of the reweighted results is much milder. To make sure that the reweighted results are in the a^2 -scaling regime, for both cases we present a non-standard approach to determine χ and compare them to reweighting.

In the first case (Extended Data Fig. 3) the temperature is just at the transition point, $T = 150$ MeV, where we expect to get a value close to the zero temperature

susceptibility. This suggests that in this case the cut-off effects of the standard method can be largely eliminated by performing the continuum limit of the ratio $\chi(T, a)/\chi(T=0, a)$, where the finite temperature result is divided by the zero temperature one at the same lattice spacing. We call this approach the ‘ratio method’, see for example, ref. 7. As can be seen in Extended Data Fig. 3, the cut-off effects are indeed reduced. The continuum extrapolation so obtained is nicely consistent with reweighting.

In the second case, Extended Data Fig. 4, we have a temperature well above the transition, $T = 300$ MeV. We see again that the standard method produces results with large cut-off effects. The ratio method seems to perform better, but the apparent scaling is misleading. Although a nice continuum extrapolation can be done from lattice spacings $N_t = 8, 10$ and 12, the $N_t = 16$ result is much below the extrapolation curve. The reweighting produces a result that is an order of magnitude smaller. Below we introduce a new method, called the ‘fixed sector integral technique’, which is tailored for large temperatures. The result so obtained, where no reweighting is applied, agrees reasonably with the reweighted one in the continuum limit.

These results provide numerical evidence for our expectations: reweighting not only produces a correct continuum limit, it also eliminates the large cut-off effects of staggered fermions.

Fixed sector integral technique. There are many proposals to increase the tunneling between the topological sectors; see, for example, refs 27–29. Here we forbid sector changes and determine the relative weight of the sectors by measuring the Q dependence of certain observables. (We note that a discussion of our method, though only in the quenched approximation using coarse lattices, has appeared in ref. 30.) Here we illustrate the method in the quenched approximation; for the extension in the case of dynamical fermions, see Supplementary Information. The gauge configurations are generated with a probability proportional to $\exp(-\beta S_g)$, where β is the gauge coupling parameter and S_g is the gauge action. We consider the following differentials:

$$b_Q \equiv - \frac{d \log Z_Q/Z_0}{d \log T} = - \frac{d\beta}{d \log a} \langle S_g \rangle_{Q=0} \quad (2)$$

where Z_Q is the partition function of the system restricted to sector Q . In the continuum limit the sectors are unambiguously defined, but on the lattice several different definitions are possible (our choice is given later on). In equation (2) we introduced the notation $\langle O \rangle_{Q=0} = \langle O \rangle_Q - \langle O \rangle_0$ for the difference of the expectation values of an observable between sectors Q and 0. Equation (2) gives a renormalized quantity, the ultraviolet divergences cancel in the difference of the gauge actions. The important observation is that the necessary statistics to reach a certain level of precision on values of b_Q does not depend on the temperature.

To obtain the relative weights Z_Q/Z_0 , we need to integrate equation (2) in the temperature. For that, we start from a temperature T_0 , where the standard approach is still feasible, and determine Z_Q/Z_0 . Then by measuring the b_Q values for higher temperatures, we can use the following integral in temperature T' to obtain Z_Q/Z_0 :

$$Z_Q/Z_0|_T = \exp \left(- \int_{T_0}^T d \log T' b_Q(T') \right) Z_Q/Z_0|_{T_0} \quad (3)$$

Let us make a remark about the volume dependence. As we increase the temperature, the ratios Z_{Q+1}/Z_Q get smaller. This effect is in competition with the infinite volume limit, which brings these ratios closer to 1. The question is how many sectors are needed to determine χ reliably. χ is an intensive quantity, and as such, its finite volume effects can be neglected, if the box size is large enough to accommodate all correlation lengths in the system. In our quenched study⁵, we found that for $LT_c \gtrsim 2$ the finite size effects on χ are negligible, where T_c denotes the phase transition temperature.

For high temperatures only the $Q = 0$ and 1 sectors remain relevant and $\langle Q^2 \rangle = \chi V$ becomes small. Using the data from our quenched simulations⁵, we found that in a box size of $L = 2/T_c$, the contribution of $Q \geq 2$ sectors to χ and also χV are on the per cent level at $1.7T_c$ and they decrease rapidly with the temperature³. (A similar behaviour was found with dynamical fermions, see Supplementary Information for details.) In the case when contributions from Z_2 and higher sectors are small it is appropriate to write $\chi = 2Z_1/Z_0(1 + 2Z_1/Z_0)/V$. To the accuracy of $\mathcal{O}(\chi V)$, one can also use $\chi \approx 2Z_1/Z_0/V$, and then the decay exponent of the susceptibility b can be simply obtained as

$$b \equiv - \frac{d \log \chi}{d \log T} \approx b_1 - 4 \quad (4)$$

Here the term -4 reflects that the physical volume also changes with the temperature. To derive the Stefan–Boltzmann limit of equation (4), we can use the fact that for large temperatures $\beta = 33 \log a/(4\pi^2)$. The gauge action difference is

given by the classical action of one instanton, the solution of the classical theory with topological charge 1, $\langle S_g \rangle_{1-0} = 4\pi^2/3$. We get $b = 7$ in the Stefan–Boltzmann limit (for $a \neq 0$ there are also some cut-off effects).

As we have already mentioned, to reach the same level of precision on b_1 as the temperature increases, the statistics can be kept constant. However, with an increase in the spatial size N_s , the statistics have to be increased as N_s^3 , and as a result the computer time goes as N_s^6 . This can be understood as follows: the gauge action difference between sectors 1 and 0 will be approximately given by the action of one instanton, which remains constant with increasing volume. The gauge action S_g itself, however, increases with the volume and the cancellation in equation (2) gets more severe. This ‘volume squared’ scaling problem can be mitigated by putting more and more topological charge into the box with increasing box size. If the topological objects are localized and well separated, then for large volumes the action difference between sectors 1 and 0 can be obtained from the difference between sectors Q and 0:

$$\langle S_g \rangle_{1-0} = \langle S_g \rangle_{Q-0} / |Q| \quad (5)$$

This relationship can be used to achieve a Q -fold increase in the signal-to-noise ratio, which translates into a Q^2 -fold decrease in the necessary computing time. Next, we are going to check this relation in our numerical simulations.

Numerical illustration. We have carried out several numerical simulations to test the new approach. Details of the algorithm and of our definition of Q can be found in the Supplementary Information.

At finite lattice spacing, Q is not necessarily an integer, thus there is a certain degree of ambiguity in defining the sectors; this ambiguity disappears in the continuum limit. First, we looked at simulations, where we constrained Q to be larger than 0.5. The parameters can be found in Extended Data Table 1 under the label ‘ N_t -scan’. The results can be seen in Extended Data Fig. 5, where the charge distributions for four different lattice spacings are plotted. Since the temperature was high, the system did not explore configurations with $Q > 1$. The non-zero width of the distributions is a lattice artefact. We can clearly observe that the peaks get sharper towards the continuum. Also, the centre of the peak gets closer to 1. This is expected, since our definition of Q , which is evaluated along the Wilson flow, is renormalized. These centre values are also given in the plot, and are denoted by z . We found that z is compatible with a $1 + c/N_t^2$ behaviour. Thus these peaks at finite lattice spacing correspond to the $Q = 1$ sector in the continuum. The z factors can be used as an $\mathcal{O}(1/N_t^2)$ correction to move the peaks of the Q distribution closer to integer values. This correction is optional for the standard evaluation of χ , but becomes useful for identifying higher Q sectors especially on coarse lattices. Inclusion of this z factor corresponds to a $\mathcal{O}(a^2)$ improvement in renormalized quantities.

Making the peaks sharper can also be achieved by using improved gauge actions, like the tree-level Symanzik, Iwasaki or DBW2 actions. They suppress the topological dislocations and produce fewer tunnelling events. For a comparison of the topological properties of these gauge actions, see ref. 31. It can also be useful to improve the definition of Q along the lines presented in ref. 32, which pushes Q towards integer values.

To explore sectors with higher Q , we defined the boundaries of the intervals as $(Q - \frac{1}{2}) \times z$. We found sharp peaks in the distribution of Q ; see, for example, Extended Data Fig. 6, where Q -histograms corresponding to the ‘ Q -scan’ simulations are shown. For the parameters, see Extended Data Table 1. The peaks are centred approximately around $Q \times z$, using the z factor found in the $Q = 1$ simulations. In these runs we went up to $Q = 8$. As can be seen, with increasing Q the charge distributions get broader. We also observe that changing the volume at this particular temperature does not have a large effect on the distribution. We note that the relative weights between the histograms are not included in the plot. These can be determined using the fixed sector integral technique.

It can also happen that a simulation gets trapped on the predefined sector boundary with a small acceptance ratio. This can be interpreted as a topological dislocation that is trying to disappear in each update, but is not allowed to disappear due to the Metropolis step. In the presented simulations, this happened in one out of 16 simulation streams, on a 8×64^3 lattice at $\beta = 7.30$ in sector $Q = 8$. In the gauge action difference the result was consistent with the untrapped streams, thus the inclusion or non-inclusion of this stream did not change the value of b_8 . Nevertheless, we discarded this stream from our final analysis, because owing to the small acceptance it had a large autocorrelation time and was obviously non-ergodic.

An important issue in fixed topology simulations is ergodicity. We used 16 streams: the starting configurations were picked from a low temperature

simulation where topology decorrelated on a timescale of few updates. Therefore the streams can be regarded as independent. After a few thousands of updates, the gauge action was consistent among the different streams. As an example, in Extended Data Fig. 7 we show the results of the Q -scan runs (see Extended Data Table 1). Plotted is b_Q from equation (2). The odd- Q sectors are not shown for clarity. The results obtained from different streams are all consistent with one another.

It is also interesting to investigate the Q dependence of b_Q . As we mentioned above, a naive expectation is that sector Q contains Q localized objects, each of them independently contributing b_1 to the total b_Q , thus $b_Q = |Q| \times b_1$ (see equation (5)). With increasing volume the corrections to this equation should get smaller, owing to the cluster decomposition principle. We found that even on 8×16^3 lattices up to $Q = 8$ the gauge action differences are consistent with a linear increase with Q . The lines in Extended Data Fig. 7 represent the fit to all streams and charges assuming equation (5). A good fit quality can be obtained. On the basis of this finding, we used the $8 - 0$ difference in a large volume run 8×64^3 , for which measuring the $1 - 0$ difference would have been much more expensive.

We now investigate the cut-off and finite size effects in b_1 at a temperature of $T \approx 6T_c$. As we have already discussed, at such a high temperature the contributions from $b_Q \geq 2$ can be safely neglected when calculating the full susceptibility, and the decay exponent is given by $b = b_1 - 4$. The upper plot in Extended Data Fig. 8 shows b as a function of the lattice spacing squared in a fixed physical volume, whereas the middle panel shows it as a function of the aspect ratio N_s/N_t . The parameters of the runs are listed in Extended Data Table 1 under the labels N_s -scan and N_t -scan. Starting from aspect ratio ~ 3 , we see no significant finite size effects. We note that starting from aspect ratio 6, the boxes are large enough to accommodate non-perturbative length scales, that is, $LT_c \geq 1$. We see no difference between boxes with perturbative and non-perturbative size.

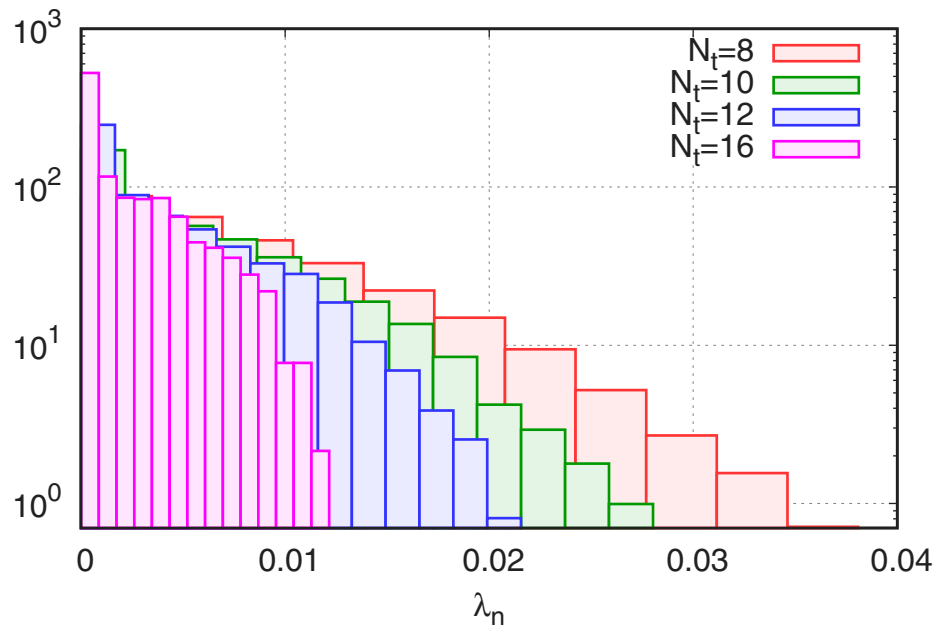
In the second set of simulations, we investigated the temperature dependence of the decay exponent. On the basis of the above results we took a lattice size of 8×32^3 ; the parameters can be found in Extended Data Table 1 under the label ‘temperature scan’. The upper panel of Extended Data Fig. 9 shows the results for $b = b_1 - 4$. Again we find agreement between the new data and the direct approach. At one temperature we did a simulation on an 8×64^3 lattice, where the exponent was obtained from measuring the difference between the $Q = 8$ and 0 sectors, $b = b_8/8 - 4$. We see no significant finite size effect.

To get the Z_1/Z_0 ratio, we performed a direct simulation at a temperature of $T_0 = 1.2T_c$. From this temperature we used a trapezoidal integration of b_1 to obtain the Z_1/Z_0 ratio as the function of temperature, up to $7T_c$. In the lower panel of Extended Data Fig. 9 we plot $\chi = 2Z_1/Z_0/(1 + 2Z_1/Z_0)/V$, normalized by T_c^4 which can be compared to the lattice result obtained from the direct method⁵. As we already discussed, starting from a temperature of $1.7T_c$ the contribution of $Q \geq 2$ can be neglected to obtain the susceptibility. We find a good agreement both for the exponent and the susceptibility itself.

Extended Data Fig. 9 was made using 30 million 8×32^3 and 1 million 8×64^3 update sweeps. The cost of a simulation at $T = 7T_c$ using the standard method can be estimated from ref. 5: it would require about 250 million updates on 8×64^3 lattices or about 2 billion on 8×32^3 , two orders of magnitude more, than with the novel method.

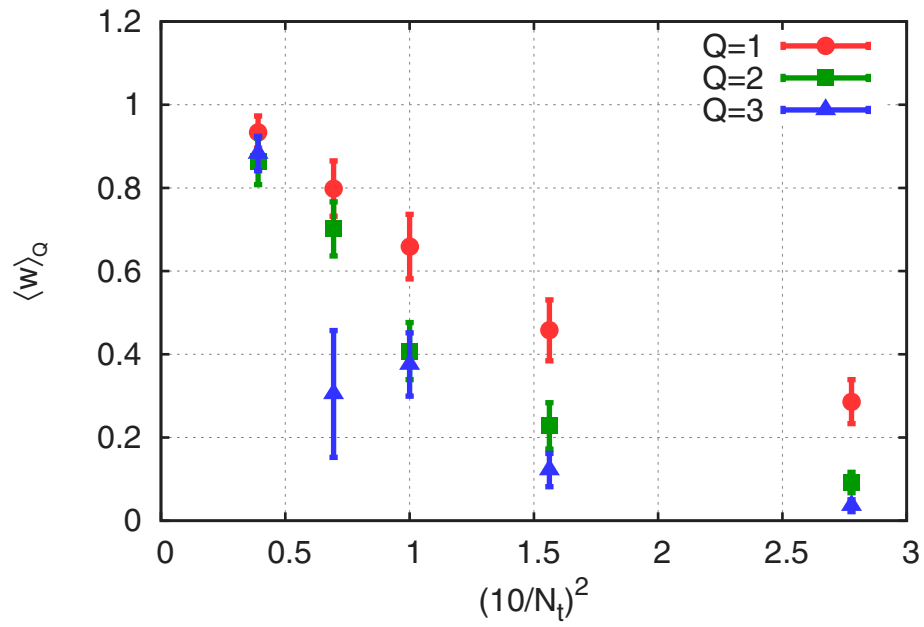
Code availability. A CPU code for configuration production can be obtained from Z.F. on request. The Wilson flow evolution code, which was used to determine Q , can be downloaded from <https://arxiv.org/abs/1203.4469>

25. Dürr, S., Hoelbling, C. & Wenger, U. Staggered eigenvalue mimicry. *Phys. Rev. D* **70**, 094502 (2004).
26. Dürr, S. Theoretical issues with staggered fermion simulations. In *Proceedings of Science* Vol. LAT2005, 021, <http://arxiv.org/abs/hep-lat/0509026> (2006).
27. Luscher, M. & Schaefer, S. Lattice QCD without topology barriers. *J. High Energy Phys.* **2011**, 036 (2011).
28. Laio, A., Martinelli, G. & Sanfilippo, F. Metadynamics surfing on topology barriers: the $CP^{(N-1)}$ case. *J. High Energy Phys.* **2016**, 89 (2015).
29. Mages, S. Lattice QCD on non-orientable manifolds. Preprint at <http://arXiv.org/abs/1512.06804> (2015).
30. Frison, J., Kitano, R., Matsufuru, H., Mori, S. & Yamada, N. Topological susceptibility at high temperature on the lattice. *J. High Energy Phys.* **2016**, 021 (2016).
31. DeGrand, T. A., Hasenfratz, A. & Kovacs, T. G. Improving the chiral properties of lattice fermions. *Phys. Rev. D* **67**, 054501 (2003).
32. DeGrand, T. A., Hasenfratz, A. & Kovacs, T. G. Topological structure in the $SU(2)$ vacuum. *Nucl. Phys. B* **505**, 417–441 (1997).
33. Ringwald, A. & Schrempf, F. Confronting instanton perturbation theory with QCD lattice results. *Phys. Lett. B* **459**, 249–258 (1999).

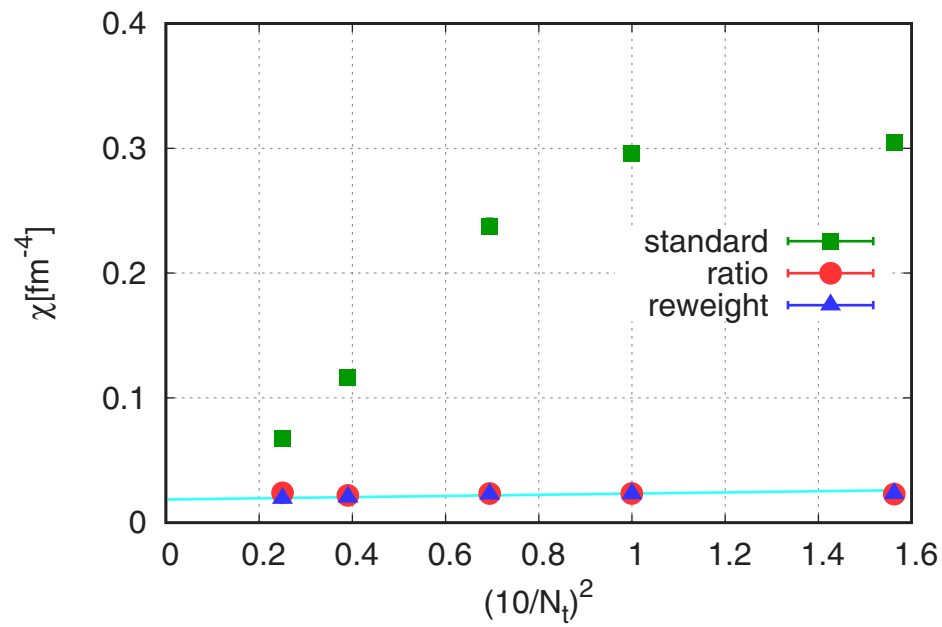


Extended Data Figure 1 | The probability distribution of the eigenvalues corresponding to the would-be zero modes. The plot shows the results of dynamical lattice QCD calculations with $n_f = 2 + 1 + 1$ flavours of staggered quarks at a temperature of $T = 240$ MeV. The result

is obtained using the chirality method described in the text. The different colours refer to different lattice spacings; see key at right. N_t is the number of lattice points in the temporal direction. Note that it is inversely proportional to the lattice spacing: $a = 1/(TN_t)$.

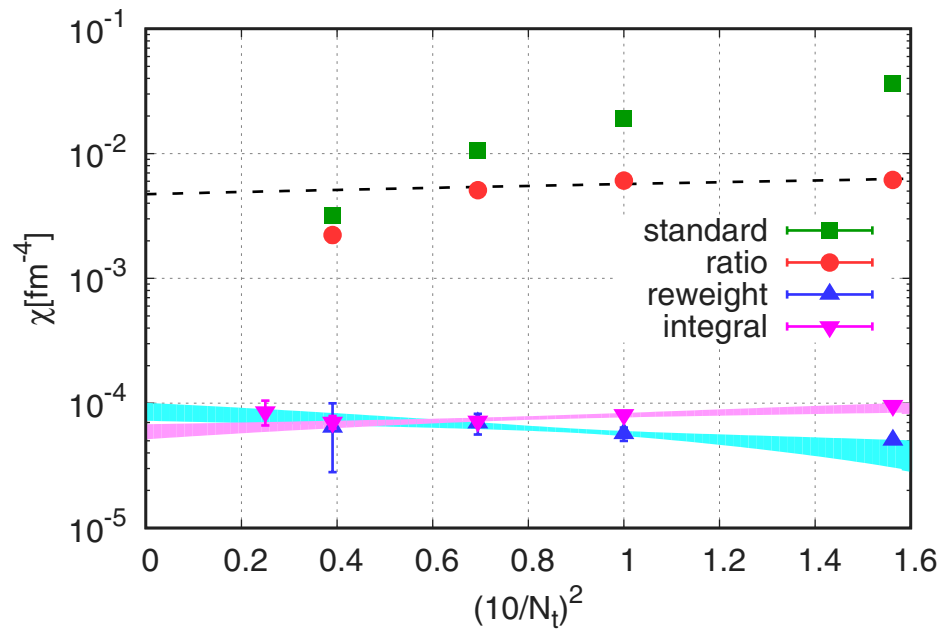


Extended Data Figure 2 | Expectation value of the weight factors in different topological sectors, $\langle w \rangle_Q$, as a function of the lattice spacing squared. The plot shows lattice results with $n_f = 3 + 1$ flavours of staggered quarks at a temperature of $T = 300$ MeV. For clarity we plot the results as a function of $10/N_t^2$. Different colours correspond to different topological charge (Q) sectors. Error bars, s.e.m.



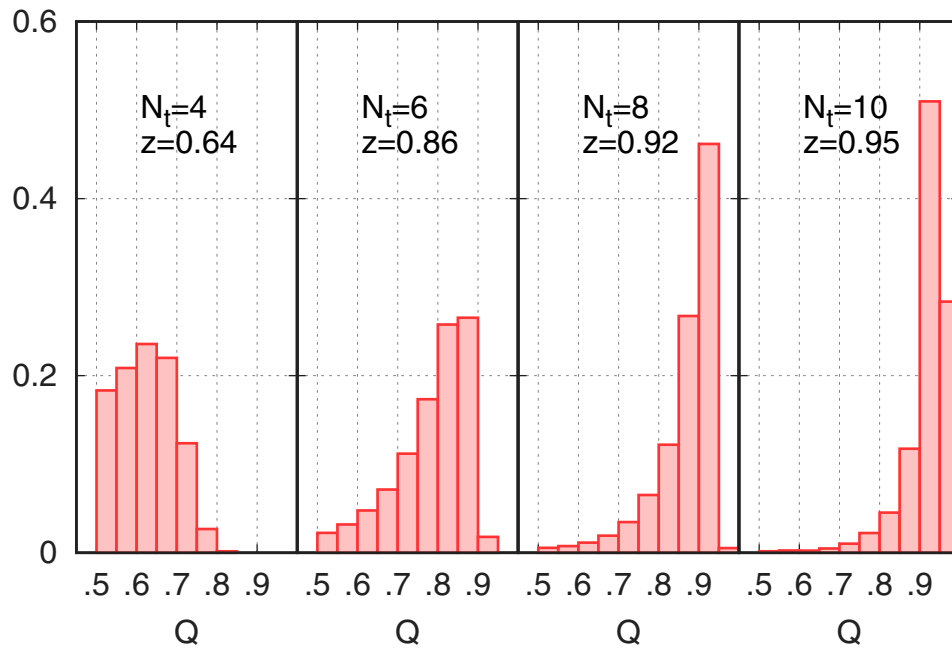
Extended Data Figure 3 | The lattice spacing dependence of the topological susceptibility χ obtained from three different methods described in the main text, namely standard, ratio and reweighting. For the reweighting method, a continuum extrapolation is also shown. The plot shows lattice results with $n_f = 2 + 1 + 1$ flavours of staggered

quarks at a temperature of $T = 150$ MeV. At this relatively low temperature the standard ('brute force') method still cannot provide three lattice spacings, which extrapolate to the proper continuum limit, though they correspond to very fine lattices with $N_t = 12, 16$ and 20 . Error bars are s.e.m. and smaller than the symbols.

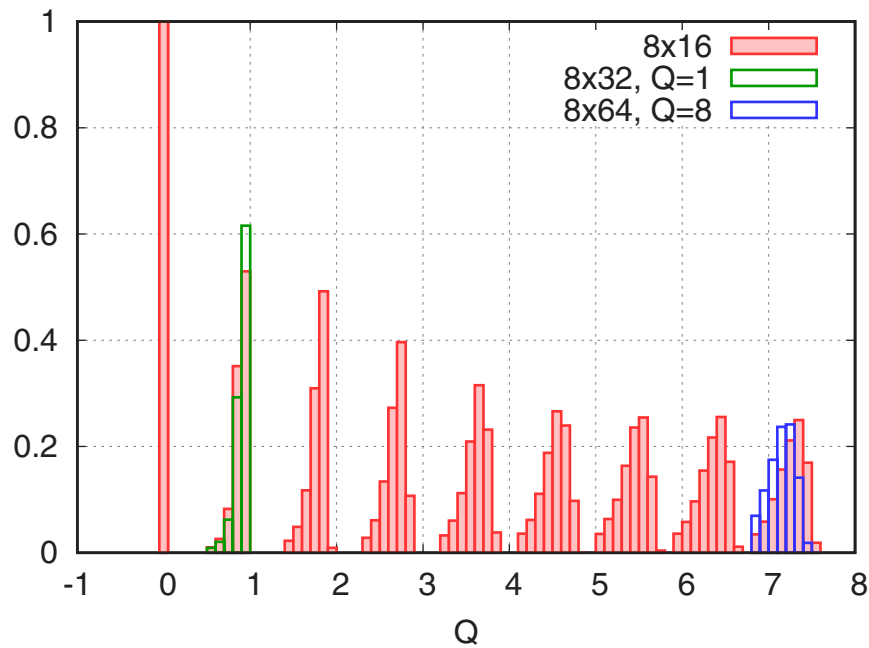


Extended Data Figure 4 | Lattice spacing dependence of the topological susceptibility obtained from four different methods described in the text, namely, standard, ratio, reweighting and integral. The plot shows lattice results with $n_f = 2 + 1 + 1$ flavours of staggered quarks

at a temperature of $T = 300$ MeV. For the ratio method, a misleading continuum extrapolation using $N_t = 8, 10$ and 12 is shown with a dashed line. For the reweighting and integral methods, continuum extrapolations are shown with bands. Error bars, s.e.m.

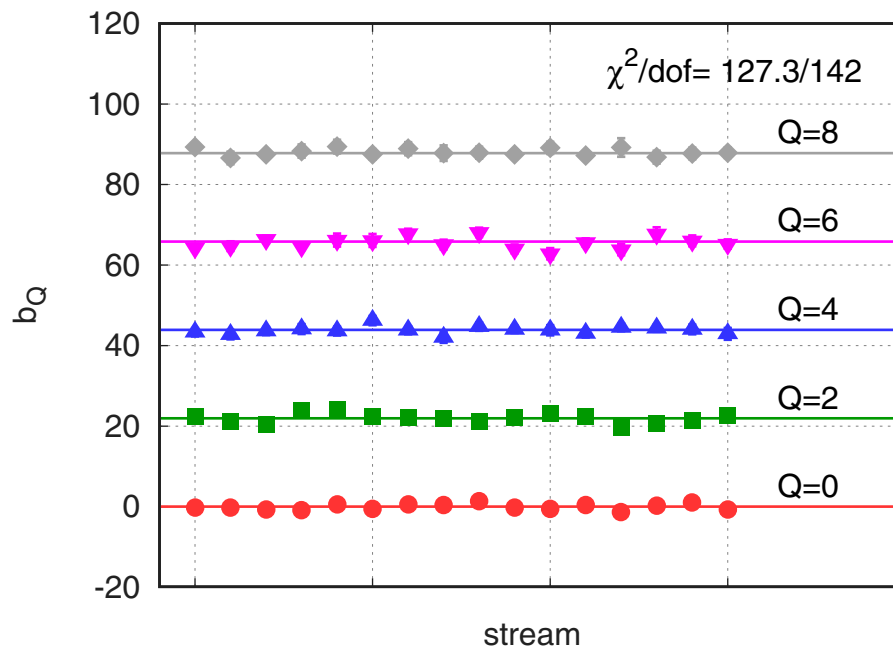


Extended Data Figure 5 | Histograms of the topological charge (Q) for different lattice spacings in simulations under the constraint $Q > 0.5$. The relative fraction of configurations in each bin is plotted. The centre of the peaks, denoted by z , is also given. The plot shows pure gauge theory simulations at $T \approx 6T_c$.



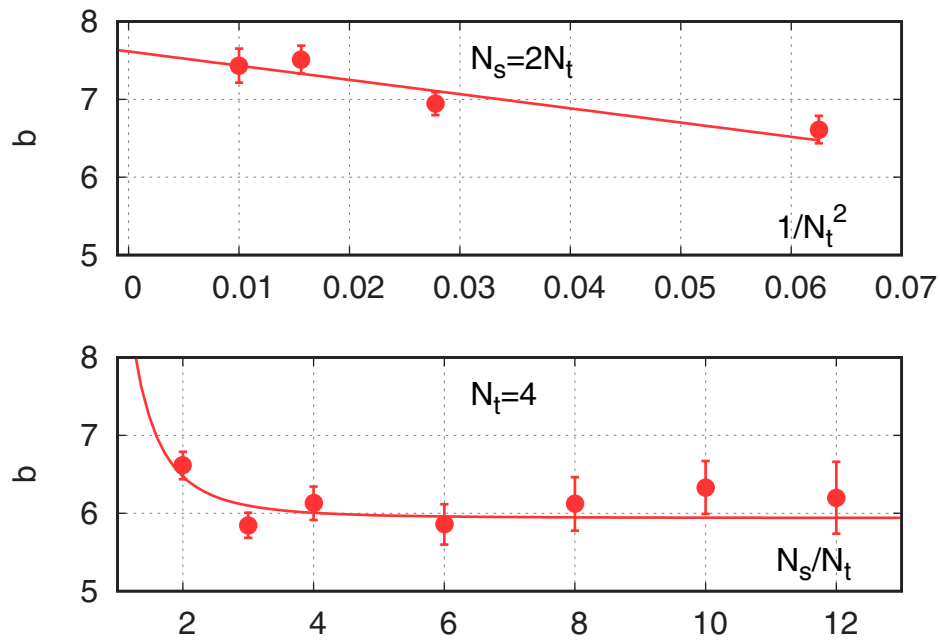
Extended Data Figure 6 | Histograms of the topological charge from fixed sector simulations for $Q = 0-8$. The relative fraction of configurations in each bin is plotted. The sector boundaries are defined using a z factor, as described in the text. Note that the relative weights between the histograms are not included in the plot. These can be

determined from the fixed sector integral technique. The plot shows pure gauge theory simulations at a temperature of $T = 5T_c$. Colour coding: red, 8×16^3 lattice with fixed topological charges $Q = 0-8$; green, 8×32^3 lattice with fixed $Q = 1$; blue, 8×64^3 lattice with fixed $Q = 8$.



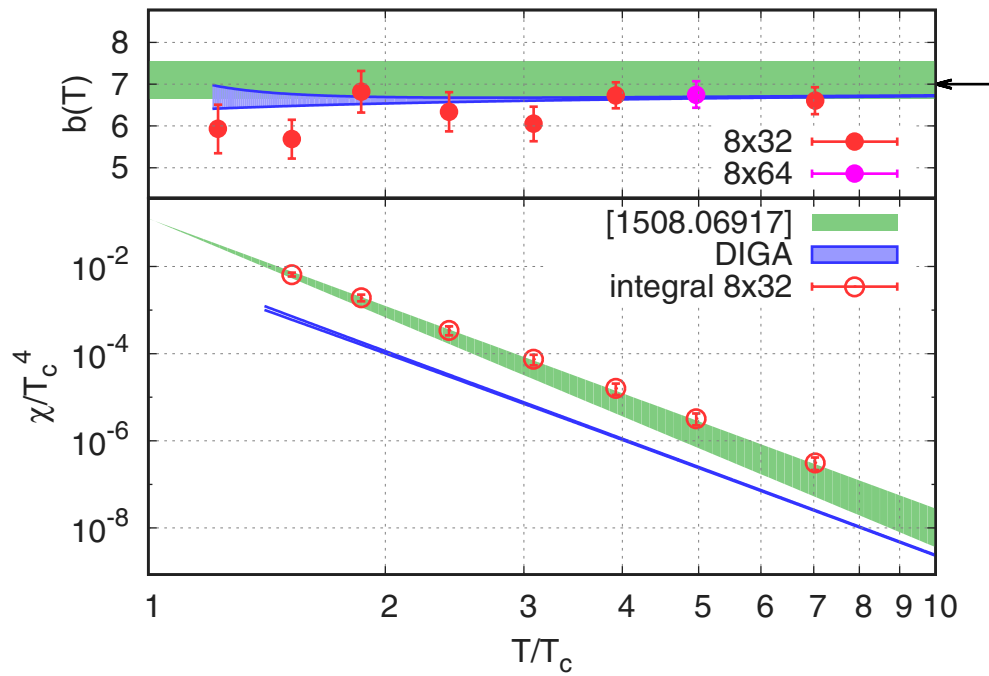
Extended Data Figure 7 | Gauge action difference. The plot shows b_Q as defined in equation (2) from pure gauge theory simulations on 8×16^3 lattices at temperature $T = 5T_c$. The different points correspond to independent simulation streams and different topological sectors. A good fit can be obtained assuming ergodicity and that the action

difference scales linearly with the topological charge, see equation (5). The horizontal lines correspond to this fit at the given Q values and the reduced χ^2 is shown at top right. Error bars are s.e.m. and smaller than the symbols.



Extended Data Figure 8 | Lattice-spacing and finite-volume dependence of the decay exponent of the topological susceptibility, b . The decay exponent b as a function of the lattice spacing squared shows the continuum extrapolation (upper panel) whereas b as a function of the linear extent of the lattice represents the infinite volume extrapolation (lower panel). The plots show pure gauge theory simulations at

temperature $T = 6T_c$. The lines are obtained from a joint fit, which takes into account both finite spacing and finite size effects. For the exponent we obtain $b = 7.1(3)$ in the continuum and infinite volume limit at this particular temperature. This is in good agreement with our previous estimate from the direct method⁵, where we obtained $b = 7.1(4)$. Error bars, s.e.m.



Extended Data Figure 9 | Topological susceptibility in the pure gauge theory. The results shown are from an earlier direct simulation³ and from the novel fixed sector integral technique. Upper panel, the temperature dependence of decay exponent b ; lower panel, temperature dependence of the topological susceptibility itself. Filled red and pink circles show the lattice results for the decay exponent on 8×32^3 and 8×64^3 lattices, respectively. The open circles show the topological susceptibility from the fixed sector integral technique. The green band refers to the result of ref. 5, the key shows the abbreviated arXiv preprint number. The black arrow indicates the Stefan–Boltzmann limit. We also show the result from the

DIGA with blue bands, see, for example, ref. 33. To convert the result into units of T_c we used $T_c/\Lambda_{\overline{MS}}=1.26$ from ref. 15. The width of the DIGA prediction reflects the change of the renormalization scale from $1/2 \pi T$ to $2 \pi T$. For the exponent b we see a good agreement for temperatures above $\sim 4T_c$, for smaller temperatures the lattice tends to give smaller values than the DIGA. In the case of the susceptibility, the DIGA underestimates the lattice result by about an order of magnitude, this has already been observed in ref. 5. The ratio at $T=2.4T_c$ is $K=11.1(2.6)$, where the error is dominated by the lattice calculation. Error bars, s.e.m.

Extended Data Table 1 | Parameters of fixed sector simulations in the pure gauge theory

β	T / T _C	N _S × N _t	M updates	Q	acc. Q ≠ 0
N _S -scan					
6.90	6.2	12 × 4	1.3	0,1	73%
		16 × 4	1.7	0,1	73%
		24 × 4	4.3	0,1	73%
		32 × 4	5.8	0,1	73%
		40 × 4	24	0,1	73%
		48 × 4	28	0,1	73%
N _t -scan					
6.90	6.2	8 × 4	0.3	0,1	72%
7.23	6.1	12 × 6	1.3	0,1	92%
7.46	6.0	16 × 8	2.9	0,1	92%
7.65	6.0	20 × 10	4.1	0,1	98%
Q-scan					
7.30	5.0	16 × 8	0.7	0	-
			0.7	1	92%
			0.7	2	88%
			0.7	3	85%
			2.6	4	83%
			2.4	5	81%
			2.4	6	78%
			2.3	7	76%
			2.3	8	73%
temperature scan					
6.20	1.2	32 × 8	3.7	0,1	88%
6.35	1.5	32 × 8	3.7	0,1	93%
6.50	1.9	32 × 8	3.7	0,1	94%
6.70	2.4	32 × 8	3.7	0,1	94%
6.90	3.1	32 × 8	3.7	0,1	94%
7.10	3.9	32 × 8	3.7	0,1	94%
7.30	5.0	32 × 8	3.7	0,1	94%
7.30	5.0	64 × 8	1.2	0,8	64%
7.60	7.0	32 × 8	3.7	0,1	94%

The first column gives the gauge coupling, the second is the temperature in units of the transition temperature T_c , the third is the lattice extent, the fourth is the collected statistics in millions of updates, the fifth shows the Q sectors used and the last column gives the acceptance rates in the $Q > 0$ sectors. The first block of parameters shows an N_s scan for which we fixed the gauge coupling and the temperature and varied the spatial extent of the lattice. The second block contains an N_t scan corresponding to a continuum extrapolation. Here the physical volume was fixed. In the Q scan we studied all topological charge sectors from $Q=0$ to $Q=8$. The last block corresponds to a temperature scan for which we fixed temporal lattice extent and changed the temperature from $1.2T_c$ up to $7T_c$. In the trivial sector we always achieved an acceptance of 100%, which simply reflects the fact that at high temperatures the probability of exploring non-trivial topologies is very small. In the $Q=1$ sector, the acceptance was about 70% on the coarsest lattice. For this we had to switch off the overrelaxation step, which makes large moves in the configuration space, and would have almost always resulted in a topology change. On finer lattices, the acceptance was around 90% or better even in the presence of overrelaxation. In the $Q > 1$ sectors, the acceptance gradually decreased as charge was increased; a simple explanation of this is that the disappearance probability of multiple instantons is approximately the sum of the individual disappearance probabilities. The worst acceptance was around 65% on an 8×64^3 lattice in sector $Q=8$.

Emergence of a turbulent cascade in a quantum gas

Nir Navon¹, Alexander L. Gaunt¹, Robert P. Smith¹ & Zoran Hadzibabic¹

A central concept in the modern understanding of turbulence is the existence of cascades of excitations from large to small length scales, or vice versa. This concept was introduced in 1941 by Kolmogorov and Obukhov^{1,2}, and such cascades have since been observed in various systems, including interplanetary plasmas³, supernovae⁴, ocean waves⁵ and financial markets⁶. Despite much progress, a quantitative understanding of turbulence remains a challenge, owing to the interplay between many length scales that makes theoretical simulations of realistic experimental conditions difficult. Here we observe the emergence of a turbulent cascade in a weakly interacting homogeneous Bose gas—a quantum fluid that can be theoretically described on all relevant length scales. We prepare a Bose–Einstein condensate in an optical box⁷, drive it out of equilibrium with an oscillating force that pumps energy into the system at the largest length scale, study its nonlinear response to the periodic drive, and observe a gradual development of a cascade characterized by an isotropic power-law distribution in momentum space. We numerically model our experiments using the Gross–Pitaevskii equation and find excellent agreement with the measurements. Our experiments establish the uniform Bose gas as a promising new medium for investigating many aspects of turbulence, including the interplay between vortex and wave turbulence, and the relative importance of quantum and classical effects.

Compared to classical fluids, superfluids present fascinating peculiarities such as irrotational and frictionless flow, which raises fundamental questions about the character of turbulent cascades^{8,9}. Numerous experiments on quantum-fluid turbulence have been performed with liquid helium, exploring both vortex^{8,10–12} and wave turbulence^{13–15}, but their theoretical understanding is hampered by the strong interactions that make first-principles descriptions intractable. The situation is *a priori* simpler for an ultracold, weakly interacting Bose gas, which is often accurately described by the Gross–Pitaevskii equation (GPE) for the complex-valued matter field $\psi(\mathbf{r}, t)$ (where $\mathbf{r} = (x, y, z)$ is the spatial position and t is time; ref. 16). This equation is widely used to model turbulence in quantum fluids^{17–21}, but the numerical results have been lacking experimental validation. Experimentally, qualitative evidence for turbulence has been seen in quantum gases^{22–25}, but quantitative comparisons with theory were hindered by the inhomogeneous density that results from harmonic trapping. Here we eliminate this problem by studying turbulence in a homogeneous quantum gas.

The basic idea of our experiment is outlined in Fig. 1. We prepare a quasi-pure Bose–Einstein condensate (BEC) of ⁸⁷Rb atoms in a cylindrical optical box⁷, and drive it out of equilibrium with a spatially uniform, oscillating force that primarily couples to the lowest, dipole-like axial mode. Our box has length $L = 27(1) \mu\text{m}$ and radius $R = 16(1) \mu\text{m}$ (here and elsewhere, errors represent 1σ uncertainties). For our typical atom number $N \approx 10^5$, the initial, equilibrium BEC has a chemical potential $\mu/k_B \approx 2 \text{ nK}$ (where k_B is the Boltzmann constant), interaction energy per particle $E_{\text{int}}/k_B \approx 1 \text{ nK}$ and negligible kinetic energy, while the critical temperature for Bose–Einstein condensation is $T_c \approx 50 \text{ nK}$. The driving force is provided by a magnetic field gradient that creates a potential $U(\mathbf{r}) = \Delta U z/L$, where the coordinate z is along the axis of

the box (Fig. 1a). The natural scale for ΔU , separating weak and strong drives, is set by μ .

Numerical simulations in Fig. 1a show the microscopic behaviour of a shaken trapped gas, which gradually changes from simple unidirectional sloshing along z to an omnidirectional turbulent flow; in addition to the wave-like motion, we observe vortex lines (depicted in red), which are detected by computing the local circulation. (Snapshots of the turbulent flow do not obey the cylindrical symmetry of the (time-dependent) Hamiltonian. In real physical systems, any such symmetry is always broken by imperfections; in our simulations the symmetry breaking is provided by the position of the numerical grid.) Here the shaking amplitude is $\Delta U/\mu = 1$ and the longest shaking time $t_s = 2 \text{ s}$ corresponds to 16 driving periods.

Experimentally, we probe the global properties of the gas by releasing it from the trap and imaging it along a radial direction (x) after a long

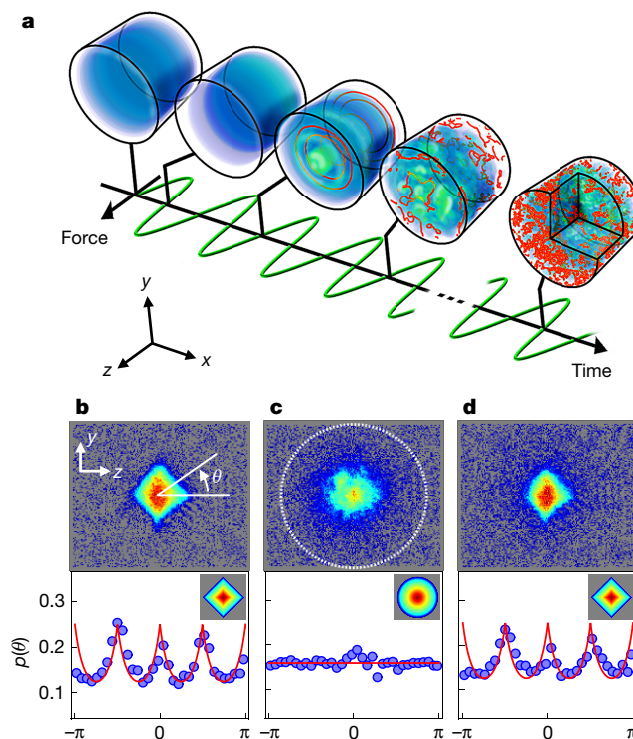


Figure 1 | From unidirectional sloshing to isotropic turbulence.

a, Gross–Pitaevskii simulations of a shaken, box-trapped Bose gas. The blue shading indicates the gas density; the red lines indicate vortices. **b–d**, Experimental absorption images taken along x after 100 ms of TOF expansion, with $N \approx 8 \times 10^4$ atoms (upper panels), and the corresponding angular distributions $p(\theta)$, averaged over 20 images taken under identical conditions (lower panels). **b**, Initial BEC; **c**, after shaking for 2 s at 8 Hz with amplitude $\Delta U/\mu \approx 1.2$; and **d**, after the turbulent cloud was allowed to relax for 1.5 s. The dashed circle in **c** corresponds to an expansion energy of $k_B T_c/2$. In the lower panels, the red lines correspond to the diamond-like and isotropic distributions depicted in the insets.

¹Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, UK.

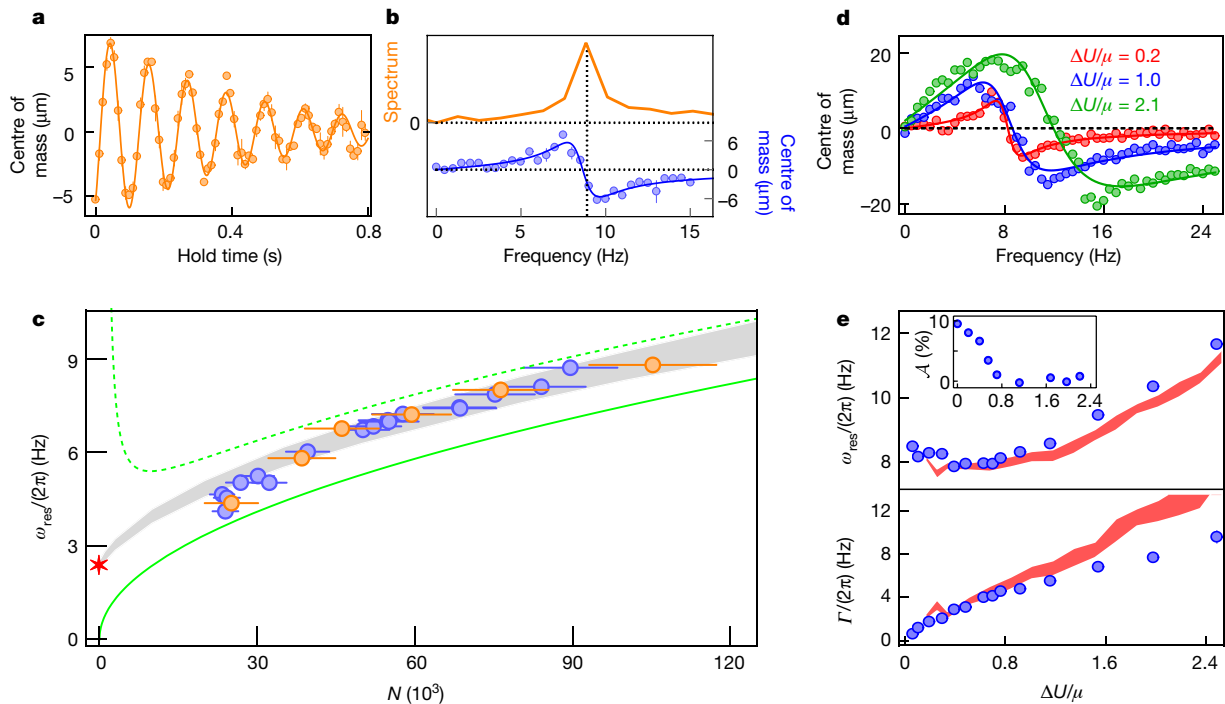


Figure 2 | Nonlinear spectroscopy of the lowest axial mode of the BEC and the route to turbulence. **a, b**, Small-amplitude oscillations of the centre of mass (along the z axis, see Fig. 1a) for $N = 10(1) \times 10^4$. **a**, Free oscillation after a 20-ms kick with $\Delta U/k_B = 1.7$ nK. **b**, Fourier spectrum of the free oscillation in **a** (orange line), and displacement after a whole number of driven oscillations, with driving amplitude $\Delta U/k_B = 0.3$ nK and $t_s = 2$ s (blue points). The solid blue line is a fit (see Methods). Error bars in **a** and **b** represent 1σ errors. **c**, Small-amplitude ω_{res} versus N for free (orange) and driven (blue) oscillations with the same ΔU and t_s values as in **a** and **b**.

time-of-flight (TOF) expansion, $t_{\text{TOF}} \geq 100$ ms. From the images we extract the centre of mass and momentum distribution of the cloud. The position of the centre of mass reflects the axial in-trap sloshing, and the evolution of the momentum distribution reveals the cascade of excitations from small to large wavevectors k —the so-called direct cascade.

In Fig. 1b–d we show a qualitative experimental signature of turbulence with three key examples of TOF images (upper panels) and the corresponding angular distributions of atoms $p(\theta)$ (lower panels). The initial BEC (Fig. 1b) shows an anisotropic expansion, which is driven by the conversion of interaction into kinetic energy, and reflects the shape of the container²⁶. In sharp contrast, after sufficiently long shaking the expansion is isotropic (Fig. 1c), with $p(\theta)$ exhibiting small fluctuations around $1/(2\pi)$. This is the first qualitative signature of a kinetic-energy-dominated turbulent state, in which the long-range coherence of the BEC is destroyed. We stress that this highly non-equilibrium state is fundamentally different from an equilibrium non-condensed state, which is also kinetic-energy-dominated and displays isotropic expansion. The key point is that, in our box trap, there is a large separation between the initial E_{int} and $k_B T_c$. This gives us access to the regime in which the total (mostly kinetic) energy per particle E satisfies $E_{\text{int}} \ll E \ll k_B T_c$. In this regime, coherence is destroyed in the turbulent state, but the corresponding equilibrium state with the same E is still deeply condensed. In Fig. 1c, the dashed circle corresponds to an expansion energy of $k_B T_c/2$, and the average energy of the atoms is clearly much lower; from the second moment of the TOF distribution we get $E \approx 0.12 k_B T_c$, which in equilibrium would correspond to a condensed fraction $\eta \approx 0.7$ (ref. 27). Indeed, if we stop shaking and allow the turbulent gas to relax, a BEC reforms (Fig. 1d) with the expected $\eta = 0.7(1)$. For all our studies of the turbulent state we restrict the shaking amplitude ($\Delta U \lesssim 2\mu$) and time

($t_s \leq 4$ s) so that $E < 0.25 k_B T_c$, which corresponds to an equilibrium η of > 0.5 . To see how pumping energy at the largest length scale (with a spatially uniform force) leads to a turbulent cascade, we perform detailed spectroscopy of the lowest-lying axial mode of the BEC (see Fig. 2). In contrast to the harmonic trap, for which the dipole mode is fixed by the trapping frequency (Kohn's theorem), in the box it depends on interactions, which results in nonlinear behaviour for non-vanishing shaking amplitudes. We first study the small-amplitude centre-of-mass response for various N , using both free and driven oscillations. In the first method, we pulse on a constant ΔU for a short time and let the gas oscillate freely in the trap for a variable hold time before releasing it and measuring the centre of mass in TOF (Fig. 2a). In the Fourier spectrum of this oscillation (orange line in Fig. 2b) we see a single strong peak, with no indication that the gradient kick directly couples to other low-lying modes. In the second method, we apply a continuous oscillating drive of (angular) frequency ω and amplitude $\Delta U \lesssim 0.5\mu$, and perform TOF measurements after a whole number of drive periods. Similarly to a driven harmonic oscillator, the displacement of the centre of mass has a dispersive line shape as a function of ω , vanishing on resonance (see Methods). As shown in Fig. 2b, the two methods give the same resonant frequency ω_{res} . In Fig. 2c we plot the small-amplitude ω_{res} versus N , and compare the data with various theories. The hydrodynamic prediction (solid green line) is $\omega_{\text{HD}} = \pi c/L$, where $c = \sqrt{\mu/m}$ is the speed of sound and m is the atom mass. This theory assumes that the healing length $\xi = \hbar/\sqrt{2m\mu}$ (where \hbar is the reduced Planck constant) satisfies $\xi \ll L$. It is therefore not applicable in the $N \rightarrow 0$ limit, where $\omega_{\text{res}} = 3\pi^2\hbar/(2mL^2)$ (red star in Fig. 2c) is given by the splitting of the lowest axial single-particle states. For our largest BEC, $L/\xi \approx 20$, but ω_{HD} is still

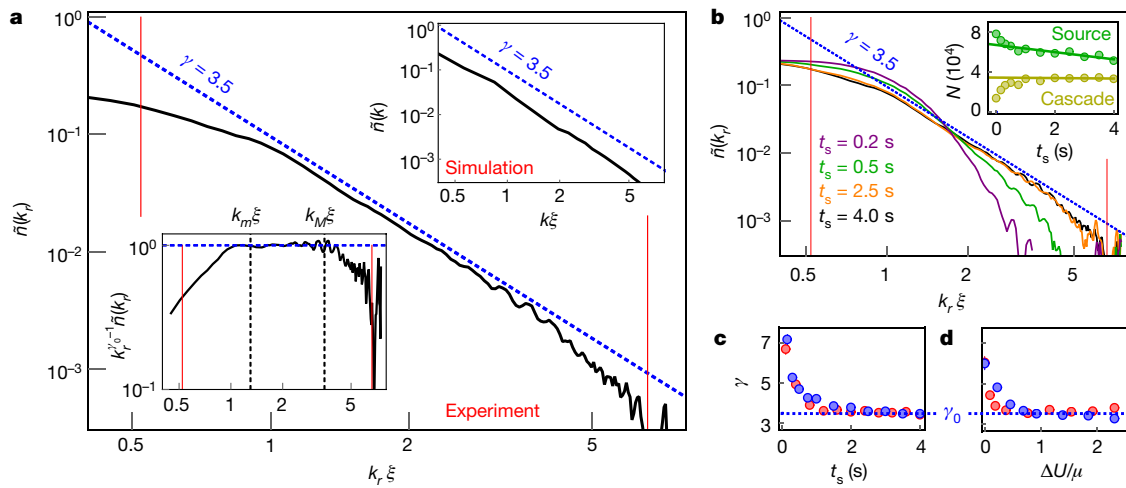


Figure 3 | Development of a turbulent cascade. **a**, Momentum distribution of the turbulent gas (solid black line), for $N = 7(1) \times 10^4$, $\Delta U/\mu = 1.1(1)$, $t_s = 4$ s, $\omega/(2\pi) = 9$ Hz and $t_{\text{TOF}} = 100$ ms. The vertical red lines indicate the momentum resolution k_{low} (left) and the energy sink at k_{high} (right); the dashed blue line is a guide to the eye, offset from the data for clarity. Lower inset, compensated spectrum $k_r^{\gamma_0-1} \tilde{n}(k_r)$ with $\gamma_0 = 3.5$ (in log–log scale); k_m and k_M define the fitting ranges used in **b–d**. Upper inset, steady-state distribution from GPE simulations, for $\Delta U/\mu = 1$. **b**, Dynamics of $\tilde{n}(k_r)$ towards the steady state, for $\Delta U/\mu = 1.1(1)$. Inset,

total atom population for $k_r < k_m$ (the low- k ‘source’; green), and for $k_m < k_r < k_M$ (in the cascade region; yellow). At long times (solid lines) $N_{\text{source}} = -3.6(1.5)$ atoms ms^{-1} , whereas $N_{\text{cascade}} = -0.2(3)$ atoms ms^{-1} is consistent with zero. All populations are corrected for losses due to the collisions with the background gas in the vacuum chamber (see Methods). **c**, Exponent γ versus shaking time in experiment (blue, $\Delta U/\mu = 1.1(1)$) and simulations (red, $\Delta U/\mu = 1$). **d**, Exponent γ versus shaking amplitude in experiment (blue) and simulations (red), for $t_s = 4$ s.

observably lower than ω_{res} . We empirically find that an upper bound on ω_{res} (dashed green line in Fig. 2c) is obtained by calculating ω_{HD} for an effective BEC volume that excludes the region within ξ of the trap walls. Finally, we linearize the GPE around the ground-state BEC solution for our box trap and numerically solve the resultant Bogoliubov equations (see Methods). These solutions are shown as the grey shaded area in Fig. 2c, which accounts for the experimental uncertainty in the box size. We find excellent agreement with the data, without any adjustable parameters.

In Fig. 2d, e we show measurements for driven oscillations with different drive strengths. Increasing ΔU shifts and broadens the resonance, and both trends are reproduced by our GPE simulations (red bands in Fig. 2e); for very large ΔU the classical-field GPE approximation may gradually break down. The line broadening, seen for any non-zero ΔU , indicates nonlinear coupling to other modes, which provides the route for the transfer of excitations into other directions and a direct cascade.

In the inset of Fig. 2e we plot the anisotropy of the TOF expansion $\mathcal{A} = (1/2) \int |p(\theta) - 1/(2\pi)| d\theta$ (see Methods) for 4 s of resonant driving. For $\Delta U \gtrsim 0.8\mu$ we observe the isotropic expansion ($\mathcal{A} \approx 0$) that qualitatively signals turbulence. A key quantitative expectation for an isotropic turbulent cascade is the emergence of a steady-state power-law momentum distribution: $n(k) \propto k^{-\gamma}$, where γ is a constant²⁸. Owing to the line-of-sight integration in absorption imaging, this corresponds to an in-plane distribution $\tilde{n}(k) \propto k^{-(\gamma-1)}$.

In Fig. 3 we present our study of $\tilde{n}(k)$ observed after a resonant drive. An isotropic expansion (from an anisotropic container) necessarily means that the in-trap kinetic energy dominates over the interaction energy, which in turn means that the TOF expansion can provide an accurate measure of the in-trap momentum distribution. Specifically, defining $k_r \equiv mr/(\hbar t_{\text{TOF}})$, where r is distance from the centre of mass in TOF, $\tilde{n}(k_r)$ should closely correspond to the in-trap $\tilde{n}(k)$ (see Methods, Extended Data Fig. 1). However, this correspondence does not hold for very low momenta ($k_r \lesssim k_{\text{low}} \equiv mL/(\hbar t_{\text{TOF}})$), owing to the convolution of the TOF distribution with the initial (in-trap) cloud shape. The highest momentum in our clouds ($k_{\text{high}} \equiv \sqrt{2mU_0}/\hbar$) is set by the trap depth $U_0 \approx k_B \times 60$ nK, which corresponds to an energy sink.

In Fig. 3a we show an example of $\tilde{n}(k_r)$, for $\Delta U/\mu = 1.1(1)$ and $t_s = 4$ s (black line in the main panel and lower inset), obtained by averaging

over 20 images and also performing an azimuthal average. Vertical red lines indicate the k_{low} and k_{high} boundaries. Away from these boundaries we observe a power-law behaviour, with $\gamma \approx 3.5$. This behaviour is even more visually evident in the lower inset, in which we plot $k_r^{\gamma_0-1} \tilde{n}(k_r)$, with $\gamma_0 \equiv 3.5$. In the top inset in Fig. 3a we show the result of GPE simulations (for $\Delta U/\mu = 1$), which also exhibit a power-law distribution. Moreover, the experiment and simulations are consistent with the same value of γ .

In Fig. 3b we present the evolution of $\tilde{n}(k_r)$ towards the turbulent steady state, as the shaking time is increased. In the inset we show (on a linear scale) the total atom populations in the low- k ‘source’ region $k_r < k_m$ and in the range $k_m < k_r < k_M$, where the power-law distribution is established in steady state (k_m and k_M are boundaries defined in the lower inset of Fig. 3a). Initially there is a net transfer of population from the source to the cascade region. The population growth in the cascade region means that the population flux through this k -space range is not constant at these early times. However, once the steady state is established, the population in the cascade k range saturates at a constant value, while the source is still slowly depleted. This is indeed what is expected for a direct cascade, in which a constant, k -independent population flux passes from the source, through the cascade range, to the high- k sink; formally, this population flux, for a given energy flux, should tend to zero as the sink is moved towards infinite energy²⁸. (For a non-infinite-energy sink, one strictly speaking has a quasi-steady state, because at very long times the source would be too depleted to support a constant-flux cascade.)

We further cross-validate our experiments and first-principles calculations by fitting the cascade exponent γ in the range $k_m < k_r < k_M$. In Fig. 3c we show that, for $\Delta U/\mu \approx 1$, the experiment and simulations exhibit very similar evolution with the shaking time, and reach a steady-state value of γ after $t_s \approx 2$ s. In Fig. 3d we plot the measured and simulated γ values versus the shaking amplitude for fixed $t_s = 4$ s. Here we see that the steady-state value of γ is essentially independent of ΔU , reinforcing the robustness of our conclusions (for small ΔU the steady state is not reached for $t_s = 4$ s; see also the inset of Fig. 2e).

We lastly discuss our findings in the context of previous theoretical work. The γ we observe in both the experiment and simulations is close to one of the scarce analytical predictions—the Kolmogorov–Zakharov direct-cascade exponent $\gamma = 3$, for the weak-wave turbulence of a

compressible superfluid²⁸. (This result is distinct from the prediction^{1,2} for the Kolmogorov energy spectrum of incompressible-flow-dominated turbulence, $\mathcal{E}(k) \propto k^{-5/3}$.) This prediction is based on an idealized theory that starts with the GPE, but neglects the role of vortices and assumes weak interactions between the waves. Our simulations show that vortices are present in the system, but the value of γ suggests that they do not play a quantitatively important part in the turbulent cascade (observed at wavenumbers $k\xi \gtrsim 1$), and the Kolmogorov–Zakharov theory is a reasonable approximation. Consistent with this, in simulations we find that, in the relevant k range, the compressible-flow contribution to the energy dominates over that of the incompressible flow (see Methods, Extended Data Fig. 2). The small difference between $\gamma = 3.5$ and the approximate $\gamma = 3$ could arise as a result of several (inter-linked) factors, such as a residual role of vortices, the non-negligible incompressible-flow energy, the fact that in reality the interactions between the waves are not necessarily weak^{15,28}, and the increasing importance of quantum pressure in the GPE with increasing k . The experimental flexibility offered by atomic gases, in particular the possibility to tune the strength of nonlinearity via Feshbach resonances, might enable better understanding of the applicability of the approximate analytical predictions, and the limitations of the classical-field methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 February; accepted 10 September 2016.

- Kolmogorov, A. N. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Dokl. Akad. Nauk SSSR* **30**, 299–303 (1941).
- Obukhov, A. On the distribution of energy in the spectrum of turbulent flow. *Dokl. Akad. Nauk SSSR* **32**, 22–24 (1941).
- Sorriso-Valvo, L. *et al.* Observation of inertial energy cascade in interplanetary space plasma. *Phys. Rev. Lett.* **99**, 115001 (2007).
- Mösta, P. *et al.* A large-scale dynamo and magnetoturbulence in rapidly rotating core-collapse supernovae. *Nature* **528**, 376–379 (2015).
- Hwang, P. A., Wang, D. W., Walsh, E. J., Krabill, W. B. & Swift, R. N. Airborne measurements of the wavenumber spectra of ocean surface waves. Part I: spectral slope and dimensionless spectral coefficient. *J. Phys. Oceanogr.* **30**, 2753–2767 (2000).
- Ghashghaie, S., Breymann, W., Peinke, J., Talkner, P. & Dodge, Y. Turbulent cascades in foreign exchange markets. *Nature* **381**, 767–770 (1996).
- Gaunt, A. L., Schmidutz, T. F., Gotlibovych, I., Smith, R. P. & Hadzibabic, Z. Bose-Einstein condensation of atoms in a uniform potential. *Phys. Rev. Lett.* **110**, 200406 (2013).
- Paoletti, M. S. & Lathrop, D. P. Quantum turbulence. *Annu. Rev. Condens. Matter Phys.* **2**, 213–234 (2011).
- Chesler, P. M., Liu, H. & Adams, A. Holographic vortex liquids and superfluid turbulence. *Science* **341**, 368–372 (2013).
- Maurer, J. & Tabeling, P. Local investigation of superfluid turbulence. *Europhys. Lett.* **43**, 29–34 (1998).
- Walmsley, P. M. & Golov, A. I. Quantum and quasiclassical types of superfluid turbulence. *Phys. Rev. Lett.* **100**, 245301 (2008).
- Bradley, D. *et al.* Direct measurement of the energy dissipated by quantum turbulence. *Nat. Phys.* **7**, 473–476 (2011).
- Ganshin, A. N., Efimov, V. B., Kolmakov, G. V., Mezhev-Deglin, L. P. & McClintock, P. V. E. Observation of an inverse energy cascade in developed acoustic turbulence in superfluid helium. *Phys. Rev. Lett.* **101**, 065303 (2008).
- Abdurakhimov, L. V., Brazhnikov, M. Y., Levchenko, A. A., Remizov, I. & Filatov, S. Turbulent capillary cascade near the edge of the inertial range on the surface of a quantum liquid. *JETP Lett.* **95**, 670–679 (2012).
- Kolmakov, G. V., McClintock, P. V. E. & Nazarenko, S. V. Wave turbulence in quantum fluids. *Proc. Natl Acad. Sci. USA* **111**, 4727–4734 (2014).
- Kagan, Y. & Svistunov, B. V. Evolution of correlation properties and appearance of broken symmetry in the process of Bose–Einstein condensation. *Phys. Rev. Lett.* **79**, 3331–3334 (1997); erratum **80**, 892 (1998).
- Nore, C., Abid, M. & Brachet, M. E. Kolmogorov turbulence in low-temperature superfluids. *Phys. Rev. Lett.* **78**, 3896–3899 (1997).
- Berloff, N. G. & Svistunov, B. V. Scenario of strongly nonequilibrated Bose–Einstein condensation. *Phys. Rev. A* **66**, 013603 (2002).
- Kobayashi, M. & Tsubota, M. Kolmogorov spectrum of superfluid turbulence: numerical analysis of the Gross–Pitaevskii equation with a small-scale dissipation. *Phys. Rev. Lett.* **94**, 065302 (2005).
- Proment, D., Nazarenko, S. & Onorato, M. Quantum turbulence cascades in the Gross–Pitaevskii model. *Phys. Rev. A* **80**, 051603 (2009).
- Reeves, M. T., Billam, T. P., Anderson, B. P. & Bradley, A. S. Inverse energy cascade in forced two-dimensional quantum turbulence. *Phys. Rev. Lett.* **110**, 104501 (2013).
- Henn, E. A. L., Sernan, J. A., Roati, G., Magalhães, K. M. F. & Bagnato, V. S. Emergence of turbulence in an oscillating Bose–Einstein condensate. *Phys. Rev. Lett.* **103**, 045301 (2009).
- Neely, T. W. *et al.* Characteristics of two-dimensional quantum turbulence in a compressible superfluid. *Phys. Rev. Lett.* **111**, 235301 (2013).
- Kwon, W. J., Moon, G., Choi, J., Seo, S. W. & Shin, Y. Relaxation of superfluid turbulence in highly oblate Bose–Einstein condensates. *Phys. Rev. A* **90**, 063627 (2014).
- Tsatsos, M. C. *et al.* Quantum turbulence in trapped atomic Bose–Einstein condensates. *Phys. Rep.* **622**, 1–52 (2016).
- Gotlibovych, I. *et al.* Observing properties of an interacting homogeneous Bose–Einstein condensate: Heisenberg-limited momentum spread, interaction energy, and free-expansion dynamics. *Phys. Rev. A* **89**, 061604 (2014).
- Schmidutz, T. F. *et al.* Quantum Joule–Thomson effect in a saturated homogeneous Bose gas. *Phys. Rev. Lett.* **112**, 040403 (2014).
- Zakharov, V. E., L’Vov, V. S. & Falkovich, G. *Kolmogorov Spectra of Turbulence* (Springer, 1992).

Acknowledgements We thank G. V. Shlyapnikov, B. Svistunov, S. Stringari, N. R. Cooper, J. Dalibard, M. J. Davis, R. J. Fletcher, M. W. Zwierlein, K. Fujimoto and M. Tsubota for discussions, and C. Eigen for experimental assistance. This work was supported by AFOSR, ARO, DARPA OLE, EPSRC (Grant No. EP/N011759/1) and ERC (QBox). The GeForce GTX TITAN X used for the numerical simulations was donated by the NVIDIA Corporation. N.N. and A.L.G. acknowledge support from Trinity College, Cambridge; R.P.S. acknowledges support from the Royal Society.

Author Contributions N.N. initiated the project, and took and analysed the data. A.L.G. wrote the code for the numerical simulations and analysed the results. Z.H. supervised the project. All authors contributed extensively to the interpretation of the data and the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.N. (nn270@cam.ac.uk).

METHODS

Experimental system. The BEC of ^{87}Rb atoms is produced in a quasi-uniform potential of a can-shaped dark optical dipole trap (see Fig. 1). The repulsive trap walls are sculpted using 532-nm laser light and a phase-imprinting spatial light modulator. They are formed by one circular tube beam (propagating along z) and two thin sheet beams that act as end caps⁷. At the end of evaporative cooling the trap depth is approximately $k_B \times 10$ nK and the condensed fraction is $\eta > 0.9$. We then slowly (over 700 ms) raise the trap depth to $U_0 \approx k_B \times 60$ nK, which does not result in any observable drop in η . Our atom number is calibrated to within 10% from the critical temperature for condensation²⁷. The gradient of the modulus of the magnetic field along z , used to shake the cloud, is calibrated by pulsing it on for a short time δt just after releasing the cloud from the trap and measuring the resulting velocity kick $\Delta U \delta t / (mL)$ in TOF.

Phase-shift measurement of the resonance. The position of the centre of mass of the cloud in TOF is $v t_{\text{TOF}}$, where v is its velocity at the time of release. In analogy with a driven damped harmonic oscillator, we assume that for a driving force proportional to $\sin(\omega t)$ in steady state $v(t) = A_\omega \cos(\omega t + \phi_\omega)$, where A_ω and ϕ_ω are the ω -dependent (in-trap) displacement amplitude and phase shift. For a shaking time t_s such that ωt_s is a multiple of 2π , the response $v(t_s) = A_\omega \cos(\phi_\omega)$ vanishes on resonance, and more generally

$$v(t_s) \propto \frac{\omega(\omega^2 - \omega_{\text{res}}^2)}{(\omega^2 - \omega_{\text{res}}^2)^2 + \Gamma^2 \omega^2} \quad (1)$$

where Γ is the damping rate. In practice we fix $t_s = 2$ s and make measurements at discrete points $\omega = j\Delta\omega$, where $\Delta\omega = 2\pi \times 0.5$ Hz and j is an integer. We then use Equation (1) to fit the data (see Fig. 2b, d) with ω_{res} and Γ as free parameters. **Numerical methods.** We implement a three-dimensional numerical simulation of the Gross–Pitaevskii equation (GPE)

$$i\hbar \frac{\partial \psi}{\partial t} = \left(-\frac{\hbar^2}{2m} \nabla^2 + V_{\text{ext}}(\mathbf{r}, t) + g |\psi|^2 \right) \psi \quad (2)$$

where the coupling constant is $g = 4\pi\hbar^2 a_s / m$, with a_s the s -wave scattering length, and $V_{\text{ext}}(\mathbf{r}, t)$ is an external potential. We have $V_{\text{ext}}(\mathbf{r}, t) = V_{\text{box}}(\mathbf{r}) + V_s(\mathbf{r}, t)$, where $V_{\text{box}}(\mathbf{r})$ is the (static) box-trap potential and $V_s(\mathbf{r}, t) = \Delta U \sin(\omega t) z / L$ is the (time-varying) shaking potential. We perform numerical simulations on a cubic grid of 256^3 points. Using a symmetrized split-step Fourier method we evolve the Bose field in time steps of $10 \mu\text{s}$ for up to 4 s. The calculations are performed at FP32 precision on an NVIDIA GeForce GTX TITAN X graphics card, and we achieve a running time of under 30 min for simulating each 1 s of dynamics.

Bogoliubov equations for the box trap. The starting point for the analysis of the linear response of the BEC to external perturbations is the GPE in Equation (2). We start by expanding $\psi(\mathbf{r}, t)$ around $\psi_0(\mathbf{r})$, the ground state in the potential $V_{\text{ext}}(\mathbf{r}, t) = V_{\text{box}}(\mathbf{r})$:

$$\psi(\mathbf{r}, t) = e^{-i\mu t / \hbar} [\psi_0(\mathbf{r}) + u(\mathbf{r})e^{-i\omega t} - v^*(\mathbf{r})e^{i\omega t}]$$

where the asterisk denotes the complex conjugate. Linearizing with respect to u and v leads to the Bogoliubov equations:

$$\begin{pmatrix} \mathcal{L} & -gn_0 \\ gn_0 & -\mathcal{L} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \hbar\omega \begin{pmatrix} u \\ v \end{pmatrix} \quad (3)$$

where

$$\mathcal{L} = -\frac{\hbar^2}{2m} \nabla^2 + V_{\text{box}}(\mathbf{r}) - \mu + 2gn_0(\mathbf{r})$$

and $n_0(\mathbf{r}) = |\psi_0(\mathbf{r})|^2$ is the ground-state density. The set of eigenvalues $\hbar\omega$ forms the spectrum of elementary excitations, and u and v give the corresponding eigenmodes. For periodic boundary conditions, n_0 is constant and the usual Bogoliubov spectrum is recovered²⁹. However, for fixed boundary conditions, the GPE cannot be solved analytically and $n_0(\mathbf{r})$ cannot be written in a simple closed form (it is not even separable in cylindrical coordinates). We solve the Bogoliubov equations by first computing $n_0(\mathbf{r})$ from an imaginary-time propagation of the GPE, and then numerically solving Equation (3). It is convenient to work in the basis of the free-particle eigenstates in a cylindrical box, so that the boundary conditions are automatically satisfied. Because we focus on the longitudinal modes, we restrict the basis to azimuthally symmetric functions. Our results for the frequency of the lowest-lying (antisymmetric) mode along z are shown in Fig. 2c as a grey shaded area (taking into account the uncertainty in the box size). In the limit of vanishing shaking amplitude, direct GPE simulations give the same results as the Bogoliubov approach.

Anisotropy analysis. To quantify the anisotropy of the TOF expansion, seen in the atomic distribution after a long t_{TOF} , we start with the column-density distribution in the TOF image $\tilde{n}(y, z) = \int n(\mathbf{r}) dx$, where $n(\mathbf{r})$ is the three-dimensional distribution and x is the imaging axis. We then define the angular density distribution:

$$p(\theta) = \frac{1}{\bar{N}} \int_{r_1}^{r_2} \tilde{n}(r \cos\theta, r \sin\theta) r dr \quad (4)$$

where the polar origin is set at the centre of mass of $\tilde{n}(y, z)$ and \bar{N} is the total atom number in the shell $[r_1, r_2]$, so that $\int_0^{2\pi} p(\theta) d\theta = 1$. An isotropic distribution corresponds to the uniform $p_{\text{iso}}(\theta) = (2\pi)^{-1}$. To quantify the anisotropy as a deviation from this uniform distribution, we introduce a simple heuristic measure:

$$\mathcal{A} = \frac{1}{2} \int_0^{2\pi} \left| p(\theta) - \frac{1}{2\pi} \right| d\theta$$

so that $\mathcal{A} = 0$ for p_{iso} and $\mathcal{A} \rightarrow 1$ for sharply-peaked distributions. For our pure BEC, the diamond-shaped TOF distribution is close to the idealized square-shaped distribution depicted in the insets of Fig. 1b, d, for which (taking $r_1 = 0$ and $r_2 \rightarrow \infty$) the angular distribution is $p(\theta) = p_0(\theta \bmod \pi/2)$, with

$$p_0(\theta) = \frac{1}{8} \begin{cases} \cos^{-2}\theta & \text{for } 0 \leq \theta < \pi/4 \\ \sin^{-2}\theta & \text{for } \pi/4 \leq \theta < \pi/2 \end{cases}$$

For this distribution, $\mathcal{A} \approx 9\%$. In experiments, any imaging imperfections lead to a positive \mathcal{A} ; for an equilibrium thermal gas ($\eta = 0$), which is known to be isotropic, we observe a small residual $\mathcal{A} \approx 3\%$, and define all our experimental values of \mathcal{A} from this baseline. For the inset of Fig. 2e, the radial integration in Equation (4) is performed in the shell defined by $r_1 = \hbar k_{\text{low}} t_{\text{TOF}} / m$ and $r_2 = \hbar k_{\text{high}} t_{\text{TOF}} / m$ (corresponding to the vertical red lines in Fig. 3a).

Measurement of the momentum distribution. The measurement of the momentum distribution using the TOF technique requires a kinetic-energy-dominated state. We assess the validity of this measurement by comparing it to Bragg spectroscopy³⁰, which can provide a faithful measurement of the momentum distribution even if the interaction energy is dominant over the kinetic energy. We shine onto the trapped cloud two off-resonant laser beams with wavevectors \mathbf{k}_1 and \mathbf{k}_2 , detuned from each other by a frequency $\Delta\nu$, such that the resultant 1D Bragg-diffraction optical lattice is aligned with the axis of the box trap, $\mathbf{k}_1 - \mathbf{k}_2 \propto \hat{z}$. The angle between the beams is such that the recoil energy of the diffracted atoms $E_r \approx k_B \times 320$ nK is larger than the trap depth ($U_0 \approx k_B \times 60$ nK), allowing the diffracted atoms to escape, and also much larger than the spread of energies in the trapped gas. Measuring the fraction of diffracted atoms as a function of $\Delta\nu$ yields the 1D momentum distribution along \hat{z} , $n_{1D}(k_z)$, which is related to the planar distribution \tilde{n} (measured in TOF) by an Abel transform:

$$n_{1D}(k_z) = \int_{-\infty}^{+\infty} \tilde{n} \left(\sqrt{k_y^2 + k_z^2} \right) dk_y$$

We use a long (20 ms) and low-power Bragg pulse to minimize Fourier broadening while always keeping the diffracted fraction below 15%. To compare the TOF and Bragg measurements we integrate our TOF images along \hat{y} .

In Extended Data Fig. 1a, we apply both methods to the initial state, a quasi-pure BEC. In this case the TOF measurement (solid green line) overestimates the width of the momentum distribution, owing to the importance of interactions during the expansion; the Bragg spectrum agrees well with the expected Heisenberg-limited distribution (dashed red line)²⁶. However, as shown in Extended Data Fig. 1b, in the relevant case of the kinetic-energy-dominated turbulent state, the Bragg and TOF measurements are in excellent agreement, validating our assumptions.

Background-gas losses. In the absence of shaking, the atom population in the trap slowly decays owing to collisions with the residual background gas in the vacuum chamber. These one-body losses are k -independent and described by an exponential decay with a time constant that we measured to be $\tau_{\text{vac}} = 13$ s. For analysing the population dynamics in the inset of Fig. 3b, we corrected all populations for this background loss by multiplying them by a common factor $\exp(t_s / \tau_{\text{vac}})$.

Numerical simulations of the turbulent cascade. In Extended Data Fig. 2a we show simulations of the dynamics of the in-plane momentum distribution in a shaken gas. Here $\tilde{n}(k)$ is computed from the spatial Fourier transform of the matter-wave field $\psi(\mathbf{r}, t_s)$. We observe that with increasing t_s the same power-law behaviour gradually extends from large to ever smaller length scales, as expected for a direct cascade.

Following the procedure outlined in ref. 17, we also study the fluid-dynamical kinetic-energy spectrum $\mathcal{E}(k)$. We start by computing $\tilde{\mathbf{w}}(\mathbf{k})$, the Fourier transform of the flow field

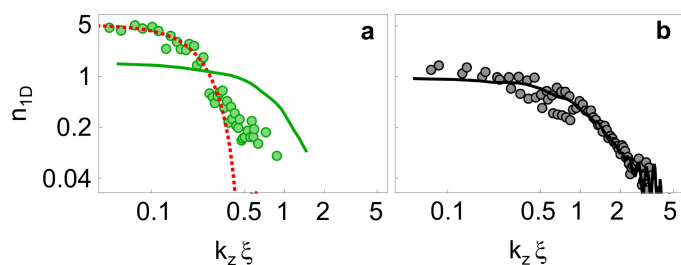
$$\mathbf{w}(\mathbf{r}) = (\hbar/m)|\psi(\mathbf{r})|\nabla\varphi(\mathbf{r})$$

where $\varphi(\mathbf{r})$ is the phase of $\psi(\mathbf{r})$ and we omit the time label for brevity. Summing $|\tilde{\mathbf{w}}(\mathbf{k})|^2$ over all momenta with $|\mathbf{k}| = k$ gives the total $\mathcal{E}(k)$. Decomposing $\tilde{\mathbf{w}}(\mathbf{k})$ into longitudinal and transverse components splits $\mathcal{E}(k)$ into the compressible- (\mathcal{E}_c) and incompressible-flow (\mathcal{E}_i) contributions, respectively. In Extended Data Fig. 2b we plot the ratio $\mathcal{E}_c(k)/\mathcal{E}_i(k)$. We find that $\mathcal{E}_c(k)$ dominates in the k range $k_m < k < k_M$, where the power-law momentum distribution is observed in both experiments

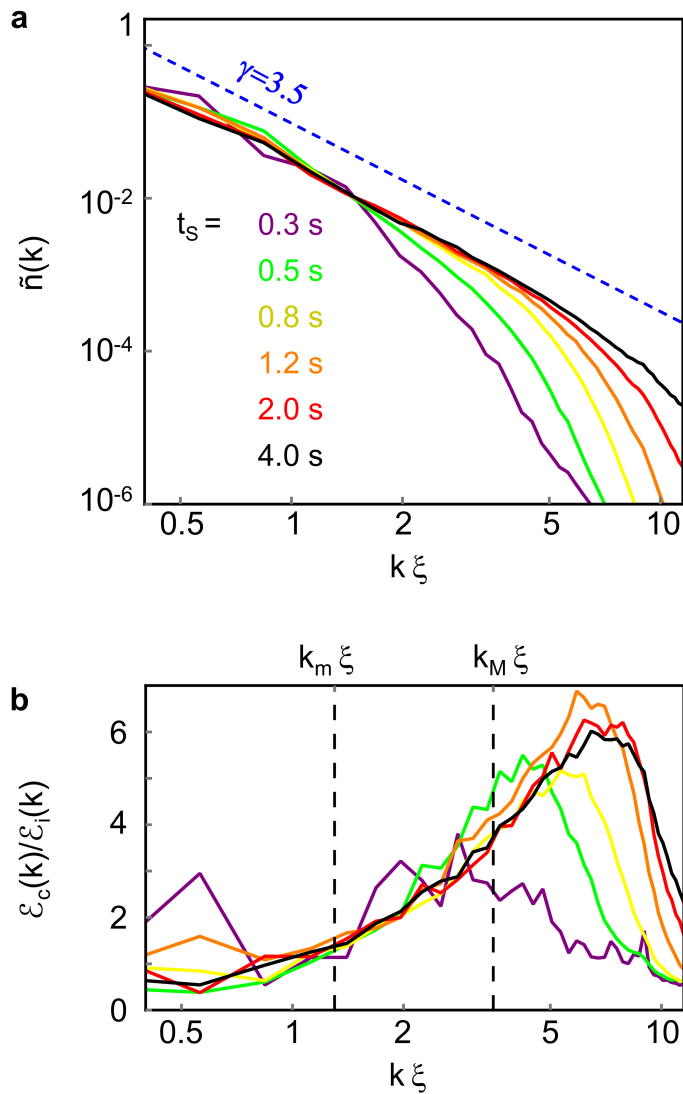
and simulations; the same numerical observation was independently made by K. Fujimoto and M. Tsubota (personal communication). This supports the view that waves play a more important part than vortices in the turbulent cascade. Also, the vortices have a core size of approximately ξ , so it qualitatively makes sense that their contribution is not important when $k \gtrsim 1/\xi$.

Data availability. Source Data for Figs 1–3 and Extended Data Figs 1, 2 are available online.

29. Pitaevskii, L. P. & Stringari, S. *Bose–Einstein Condensation and Superfluidity* Ch. 12 (Oxford Univ. Press, 2016).
30. Stenger, J. *et al.* Bragg spectroscopy of a Bose–Einstein condensate. *Phys. Rev. Lett.* **82**, 4569–4573 (1999).



Extended Data Figure 1 | Momentum distributions from TOF and Bragg techniques. **a, b,** Comparison of $n_{1D}(k_z)$ obtained using TOF expansion (solid lines) and Bragg spectroscopy (points), in the case of the initial, quasi-pure BEC (**a**) and the turbulent gas (**b**). The red dashed line in **a** corresponds to the Heisenberg-limited momentum distribution. All distributions are normalized to unity ($\int_0^\infty n_{1D} d(k_z \xi) = 1$), without any adjustable parameters.



Extended Data Figure 2 | Turbulent cascade in numerical simulations.
a, In-plane momentum distribution $\tilde{n}(k)$ for various shaking times t_s .
b, Ratio of the compressible- (\mathcal{E}_c) to incompressible-flow (\mathcal{E}_i) components of the fluid-dynamical kinetic energy, with the colours corresponding to the shaking times in **a**. The simulation parameters for both panels are $N=8 \times 10^4$, shaking frequency $\omega/(2\pi)=9 \text{ Hz}$ and shaking amplitude $\Delta U=\mu$.

Metal–organic frameworks as selectivity regulators for hydrogenation reactions

Meiting Zhao^{1*}, Kuo Yuan^{1,2*}, Yun Wang³, Guodong Li¹, Jun Guo¹, Lin Gu⁴, Wenping Hu^{2,5}, Huijun Zhao³ & Zhiyong Tang¹

Owing to the limited availability of natural sources, the widespread demand of the flavouring, perfume and pharmaceutical industries for unsaturated alcohols is met by producing them from α,β -unsaturated aldehydes, through the selective hydrogenation of the carbon–oxygen group (in preference to the carbon–carbon group)¹. However, developing effective catalysts for this transformation is challenging^{2–7}, because hydrogenation of the carbon–carbon group is thermodynamically favoured⁸. This difficulty is particularly relevant for one major category of heterogeneous catalyst: metal nanoparticles supported on metal oxides. These systems are generally incapable of significantly enhancing the selectivity towards thermodynamically unfavoured reactions, because only the edges of nanoparticles that are in direct contact with the metal–oxide support possess selective catalytic properties; most of the exposed nanoparticle surfaces do not^{9–14}. This has inspired the use of metal–organic frameworks (MOFs) to encapsulate metal nanoparticles within their layers or inside their channels, to influence the activity of the entire nanoparticle surface while maintaining efficient reactant and product transport owing to the porous nature of the material^{15–18}. Here we show that MOFs can also serve as effective selectivity regulators for the hydrogenation of α,β -unsaturated aldehydes. Sandwiching platinum nanoparticles between an inner core and an outer shell composed of an MOF with metal nodes of Fe^{3+} , Cr^{3+} or both (known as MIL-101; refs 19–21) results in stable catalysts that convert a range of α,β -unsaturated aldehydes with high efficiency and with significantly enhanced selectivity towards unsaturated alcohols. Calculations reveal that preferential interaction of MOF metal sites with the carbon–oxygen rather than the carbon–carbon group renders hydrogenation of the former by the embedded platinum nanoparticles a thermodynamically favoured reaction. We anticipate that our basic design strategy will allow the development of other selective heterogeneous catalysts for important yet challenging transformations.

The coordinatively unsaturated metal sites (CUSs) inside MOFs can be readily tuned to adjust interactions between MOFs and reactants, to activate targeted chemical bonds in reactants and thereby to lower the reaction-energy barrier of the desired chemical transformation²². We explore this concept with MOF sandwich nanostructures that are then used as hydrogenation catalysts (Supplementary Fig. 1). The nanostructures contain a layer of platinum (Pt) nanoparticles encapsulated between a core and a shell made of MIL-101, which contains either Fe^{3+} or Cr^{3+} trimers, connected with 1,4-benzenedicarboxylate (BDC) linkers (Fig. 1a and Supplementary Note 1)^{19–21}. All synthesized products are of uniform octahedral shape, with 2.8-nm Pt nanoparticles immobilized between the MIL-101 core and shell (Fig. 1b–l and Supplementary Figs 2–13). Catalytic performance is evaluated using five typical sandwich nanostructures: two MIL-101(Fe)@Pt@MIL-101(Fe) systems with shell thicknesses of about 9.2 nm and 22.0 nm (denoted as MIL-101(Fe)@Pt@MIL-101(Fe)^{9,2} and MIL-101(Fe)@

Pt@MIL-101(Fe)^{22,0}; Fig. 1b–e and Supplementary Figs 8, 9); one MIL-101(Cr)@Pt@MIL-101(Cr) with a shell thickness of around 5.1 nm (denoted as MIL-101(Cr)@Pt@MIL-101(Cr)^{5,1}; Fig. 1f, g and Supplementary Fig. 10); and two MIL-101(Cr)@Pt@MIL-101(Fe) systems with shell thicknesses of around 2.9 nm and 8.8 nm (denoted as MIL-101(Cr)@Pt@MIL-101(Fe)^{2,9} and MIL-101(Cr)@Pt@MIL-101(Fe)^{8,8}; Fig. 1h–l, Supplementary Figs 11–13 and Supplementary Note 2). All of the materials contain single crystalline MIL-101 with a face-centred-cubic (fcc) structure, and a lattice fringe of 2.6 nm that corresponds to the (222) planes of MIL-101 (ref. 19; Fig. 1c–k and Supplementary Figs 14–16). The embedded Pt nanoparticles exhibit high crystallinity, with an interplanar spacing of 0.23 nm that corresponds to the (111) planes of fcc platinum (Fig. 1e, k, insets).

Table 1 summarizes the results obtained when using the sandwich MIL-101@Pt@MIL-101 nanostructures as catalysts for the selective hydrogenation of cinnamaldehyde (A) to cinnamyl alcohol (B), along with results for Pt nanoparticles, MIL-101 and MIL-101@Pt as controls (see also Fig. 1, Supplementary Figs 2–7, and Supplementary Tables 1, 2). Because the C=O and C=C groups in A are both possible hydrogenation targets, we expect a product mixture containing the targeted cinnamyl alcohol (B), as well as hydrocinnamaldehyde (C) and phenyl propanol (D). For a more meaningful evaluation of selectivity for the desired product B, we first compared the catalysts' performances at the same conversion level (about 45%) of A. We found no noticeable hydrogenation of A by MIL-101 (Table 1, entry 9), whereas Pt nanoparticles catalyse this reaction with a turnover frequency (TOF) of 372.4 h^{-1} and a selectivity for B of only around 18.3% (Table 1, entry 8), consistent with density functional theory (DFT) predictions (Supplementary Fig. 17).

The third controls—the MIL-101@Pt nanostructures—exhibit dramatically increased selectivity for B (Table 1, entries 6 and 7), with 86.4% for MIL-101(Fe)@Pt and 44.0% for MIL-101(Cr)@Pt. This remarkable promotion of selective hydrogenation of the C=O bond by MIL-101 is probably due to the CUSs in MIL-101 acting as Lewis acid sites that interact with the C=O bond and activate it²³. A Fourier transform infrared (FTIR) survey shows an obvious redshift of the $\nu_{\text{C=O}}$ bond of A after mixing with MIL-101, while the $\nu_{\text{C=C}}$ bond remains unaltered, confirming a selective interaction between the C=O bond of A and MIL-101 (Supplementary Fig. 18).

DFT calculations with consideration of spin-polarization effect—aimed at understanding the origin of the interaction between MIL-101 and A—show that the five-coordinated metal nodes in the $\text{Fe}_3\text{OCl}(\text{COO})_6\text{H}_2\text{O}$ or $\text{Cr}_3\text{OCl}(\text{COO})_6\text{H}_2\text{O}$ trimers of supertetrahedral MIL-101 cells serve as active sites that interact with the C=O group of A (Supplementary Fig. 19)¹⁹. The calculated adsorption energy of A over the $\text{Fe}_3\text{OCl}(\text{COO})_6\text{H}_2\text{O}$ and $\text{Cr}_3\text{OCl}(\text{COO})_6\text{H}_2\text{O}$ trimers is -1.26 eV and -1.01 eV , respectively (Fig. 2); that is, binding between the trimers and A through Fe–O or Cr–O interactions is thermodynamically favoured, with the interaction between A and

¹Chinese Academy of Sciences (CAS) Key Laboratory of Nanosystem and Hierarchy Fabrication, CAS Center for Excellence in Nanoscience, National Center for Nanoscience and Technology, Beijing 100190, China. ²School of Science, Tianjin University, Tianjin 300072, China. ³Centre for Clean Environment and Energy, Gold Coast Campus, Griffith University, Queensland 4222, Australia. ⁴Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. ⁵Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China.

*These authors contributed equally to this work.

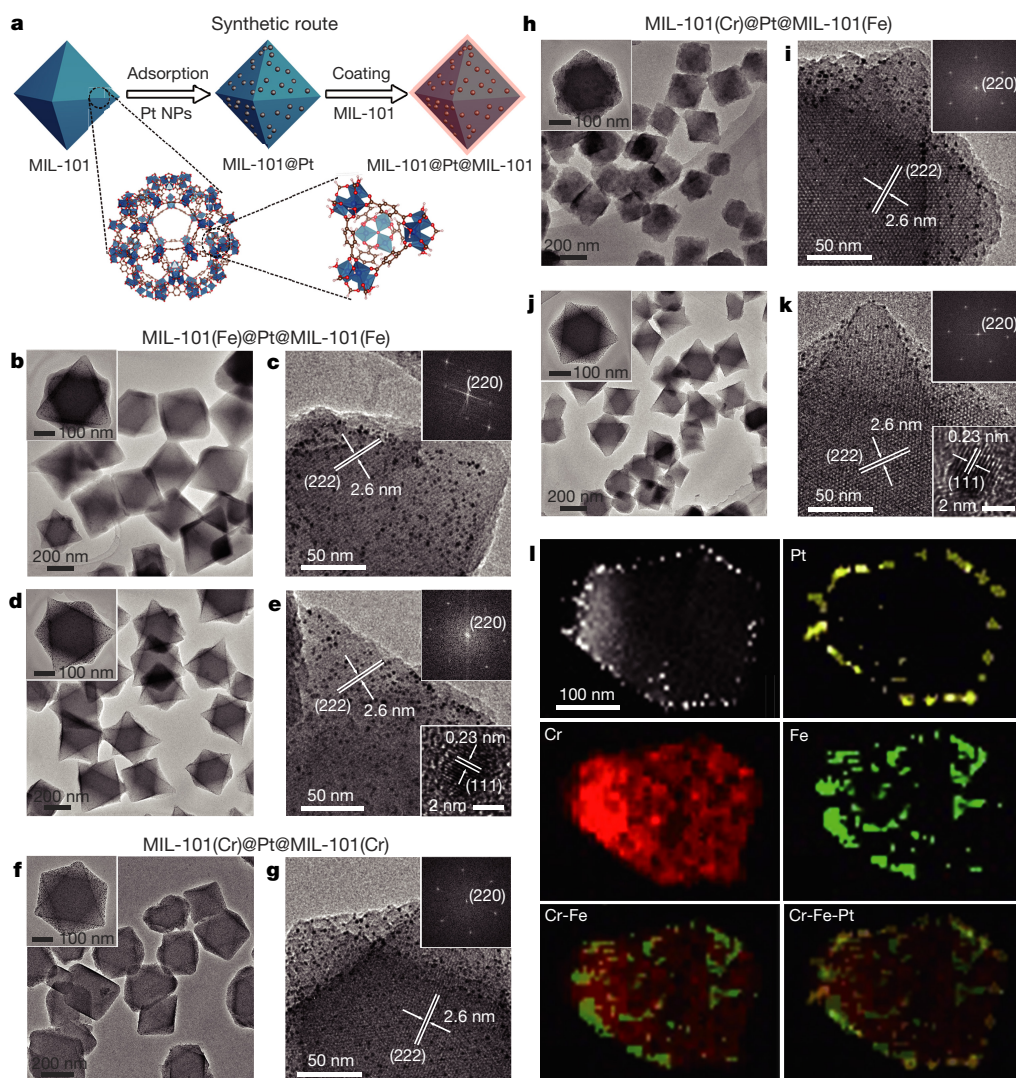


Figure 1 | Sandwich MIL-101@Pt@MIL-101 nanostructures. **a**, Synthetic route to generating sandwich MIL-101@Pt@MIL-101, comprising Pt nanoparticles (NPs) sandwiched between a core and a shell of MIL-101. **b, d**, Transmission electron microscopy (TEM) images of MIL-101(Fe)@Pt@MIL-101(Fe)^{22.0} (**b**) and MIL-101(Fe)@Pt@MIL-101(Fe)^{9.2} (**d**). **c, e**, High-resolution TEM (HR-TEM) images of MIL-101(Fe)@Pt@MIL-101(Fe)^{22.0} (**c**) and MIL-101(Fe)@Pt@MIL-101(Fe)^{9.2} (**e**). **f**, TEM image of MIL-101(Cr)@Pt@MIL-101(Cr)^{5.1}. **g**, HR-TEM image of MIL-101(Cr)@Pt@MIL-101(Cr)^{5.1}. **h, j**, TEM images of MIL-101(Cr)@Pt@MIL-101(Fe)^{8.8} (**h**) and MIL-101(Cr)@Pt@MIL-101(Fe)^{2.9} (**j**). **i, k**, HR-TEM

images of MIL-101(Cr)@Pt@MIL-101(Fe)^{8.8} (**i**) and MIL-101(Fe)@Pt@MIL-101(Fe)^{2.9} (**k**). Insets in **b, d, f, h, j**, TEM images of a single MIL-101@Pt@MIL-101 nanostructure. Insets in **c, e, g, i, k**, corresponding fast Fourier transform (FFT) images of MIL-101@Pt@MIL-101 (top right) and HR-TEM images of Pt NPs (bottom right). **l**, High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image (top left) and corresponding energy-dispersive X-ray spectroscopy (EDS) elemental mapping images of a cross-section of single MIL-101(Cr)@Pt@MIL-101(Fe)^{2.9}.

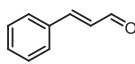
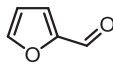
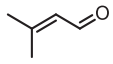

Fe₃OCl(COO)₆H₂O being the stronger one (Supplementary Figs 20–22 and Supplementary Note 3). The calculated reaction energies for hydrogenation of **A** to **B** or **C** over MIL-101(Fe)@Pt are -2.15 eV and -1.89 eV (Fig. 2a), indicating that the formation of **B** (selectivity approximately 86.4%) is energetically favoured over that of **C** (selectivity roughly 9.6%). In the case of MIL-101(Cr)@Pt, the corresponding reaction energies for **B** and **C** formation are -0.69 eV and -0.92 eV, respectively (Fig. 2b), indicating a smaller difference in the selectivity of **B** (roughly 44.0%) over **C** (about 40.0%). These results confirm the experimental observation that the CUSs in MIL-101 preferentially interact with the C=O group (rather than the C=C group) of **A** and activate it, giving rise to the improved selectivity for the formation of product **B**.

We also note that at a conversion level of 45%, TOF values are 122.1 h^{-1} for MIL-101(Fe)@Pt and 372.4 h^{-1} for MIL-101(Cr)@Pt (Table 1, entries 6 and 7), indicating that MIL-101(Cr) does not affect

the catalytic activity of Pt nanoparticles whereas MIL-101(Fe) causes a considerable decrease. X-ray photoelectron spectroscopy (XPS) measurements reveal a partial transfer of electrons from the Pt nanoparticles to MIL-101(Fe)²⁴ (Fig. 3a, b, d and Supplementary Note 4), whereas there is no obvious electron transfer between from the Pt nanoparticles to MIL-101(Cr) (Fig. 3a, c, d). This suggests the reduced electron density of Pt nanoparticles in MIL-101(Fe)@Pt as a possible cause of the decreased hydrogenation activity. This is further supported by DFT calculations showing that the negative binding energy of hydrogen atoms (ΔE_{H}) on a platinum(111) (2×2) surface increases with positive charging of the platinum surface (Supplementary Fig. 23), with the resulting stronger surface binding of the hydrogen atoms reducing the catalytic activity of MIL-101(Fe)@Pt²⁵.

Guided by these observations, and given that MIL-101 possesses two characteristic pore windows (of about 1.3 nm and 1.5 nm) that are large enough to accommodate **A** ($1.05 \text{ nm} \times 0.65 \text{ nm}$; Supplementary

Table 1 | Selective hydrogenation of different α,β -unsaturated aldehydes by different catalysts

Reaction							
$\text{R}-\text{CH}=\text{CH}-\text{CHO} \xrightarrow{\text{H}_2} \text{R}-\text{CH}=\text{CH}-\text{OH} + \text{R}-\text{CH}_2-\text{CHO} + \text{R}-\text{CH}_2-\text{CH}_2-\text{OH}$							
A	B	C	D				
Cinnamaldehyde	Furfural	3-methyl-2-butenal	Acrolein				
							
Entry	Catalysts	Time (h)	Conversion (%) [†]	Selectivity (%) [†]			TOF [‡] (h ^{−1})
				B	C	D	
Substrate: cinnamaldehyde*							
1	MIL-101(Fe)@Pt@MIL-101(Fe) ^{9,2}	8.5	45.0	94.1	2.9	3.0	18.0
2	MIL-101(Fe)@Pt@MIL-101(Fe) ^{22,0}	10	45.0	96.3	1.9	1.8	15.3
3	MIL-101(Cr)@Pt@MIL-101(Fe) ^{2,9}	2	45.0	94.2	4.7	1.1	76.3
4	MIL-101(Cr)@Pt@MIL-101(Fe) ^{8,8}	6	45.0	94.5	2.8	2.7	25.4
5	MIL-101(Cr)@Pt@MIL-101(Cr) ^{5,1}	0.7	45.0	79.2	14.6	6.2	218.1
6	MIL-101(Fe)@Pt	1.25	45.0	86.4	9.6	4.0	122.1
7	MIL-101(Cr)@Pt	0.41	45.0	44.0	40.0	16.0	372.4
8	Pt NPs	0.41	45.0	18.3	60.6	21.1	372.4
9	MIL-101(Cr) or MIL-101(Fe)	24	0	—	—	—	0
10	MIL-101(Fe)@Pt@MIL-101(Fe) ^{9,2}	24	94.3	97.0	0	3.0	13.3
11	MIL-101(Fe)@Pt@MIL-101(Fe) ^{22,0}	24	86.3	97.4	0.3	2.3	12.2
12	MIL-101(Cr)@Pt@MIL-101(Fe) ^{2,9}	20	99.8	95.6	0.8	3.6	16.9
13	MIL-101(Cr)@Pt@MIL-101(Fe) ^{8,8}	20	90.6	94.3	2.1	3.6	15.4
14	MIL-101(Cr)@Pt@MIL-101(Cr) ^{5,1}	2	95.1	62.6	12.2	25.2	161.3
Substrate: furfural*							
15	MIL-101(Fe)@Pt@MIL-101(Fe) ^{9,2}	15	85.6	93.2	3.5	3.3	19.4
16	MIL-101(Cr)@Pt@MIL-101(Fe) ^{8,8}	7	87.9	96.5	3.5	0	42.6
17	MIL-101(Cr)@Pt@MIL-101(Cr) ^{5,1}	5	98.5	99.8	0	0.2	66.8
Substrate: 3-methyl-2-butenal*							
18	MIL-101(Fe)@Pt@MIL-101(Fe) ^{9,2}	17	71.1	87.8	3.2	9.0	14.2
19	MIL-101(Fe)@Pt@MIL-101(Fe) ^{22,0}	24	59.9	92.5	2.8	4.7	8.5
20	MIL-101(Cr)@Pt@MIL-101(Fe) ^{8,8}	10	85.0	64.4	7.9	27.7	28.8
Substrate: acrolein*							
21	MIL-101(Fe)@Pt@MIL-101(Fe) ^{9,2}	3	76.9	75.7	19.0	5.3	86.9
22	MIL-101(Fe)@Pt@MIL-101(Fe) ^{22,0}	3	52.7	97.3	2.7	0	59.6
23	MIL-101(Cr)@Pt@MIL-101(Fe) ^{8,8}	0.75	87.5	57.9	17.3	24.8	395.8
24	Au ₂₅ (SR) ₁₈ /Fe ₂ O ₃ [¶]	3	47.0	92.0	—	—	3.1
25	Au clusters coated with <i>tert</i> -butyl (naphthalen-1-yl)phosphine oxide [§]	18	61.0	>99	—	—	1.7

A, α,β -unsaturated aldehydes (cinnamaldehyde, furfural, 3-methyl-2-butenal or acrolein); **B**, unsaturated alcohol; **C**, saturated aldehyde; **D**, saturated alcohol.

*Reaction condition: each catalyst contains 0.23 mg Pt NPs; the reaction requires 0.4 mmol of **A**, room temperature and 3 MPa H₂.

[†]The percentage conversion of **A** and selectivity for specific products were determined by gas chromatography-mass spectrometry and gas chromatography or ¹H NMR.

[‡]TOF was calculated by the mole number of converted **A** (mole number of total Pt or Au)^{−1} h^{−1}.

[¶]Reaction condition²⁶: 100 mg 1%Au₂₅(SR)₁₈/Fe₂O₃, 0.1 mmol **A**, 0 °C and 0.1 MPa H₂. The catalytic selective hydrogenation of acrolein has been described previously.

[§]Reaction condition²⁷: 0.01 mmol Au clusters (assuming 46.28% Au on the basis of element analysis), 0.5 mmol **A**, 5 ml tetrahydrofuran, 40 °C and 4 MPa H₂. The catalytic selective hydrogenation of acrolein has been described previously.

Table 2 and Figs 24, 25), we next explored the sandwich MIL-101@Pt@MIL-101 structure as an ideal catalyst for selective hydrogenation of **A** to **B**. Entries 1–5 in Table 1 confirm that coating with MIL-101 shells of different thicknesses considerably enhances the selectivity for **B**: the selectivity with MIL-101(Cr)@Pt@MIL-101(Cr)^{5,1} is 79.2%, while using MIL-101(Cr) or MIL-101(Fe) cores that are coated with MIL-101(Fe) shells gives selectivities that are always higher than 94%. The improved selectivities are accompanied by a decrease in TOF values, which is expected because the MIL-101 shells slow down reactant and product diffusion. Coating with MIL-101(Fe) leads to an additional decrease in catalytic activity, owing to the interfacial electron-transfer effect (Fig. 3 and Supplementary Fig. 23).

Our MIL-101@Pt@MIL-101 catalysts combine exceptionally high selectivity and conversion efficiency (Table 1, entries 10–14, and Supplementary Table 3), and outperform state-of-the-art catalysts (Supplementary Table 4). MIL-101(Cr)@Pt@MIL-101(Fe)^{2,9}, for example, exhibits excellent selectivity (95.6%) and almost full conversion

(99.8%) (Supplementary Fig. 26). Moreover, reusability tests indicate that for MIL-101(Cr)@Pt@MIL-101(Fe)^{2,9}, both the conversion efficiency of **A** and the selectivity for **B** remain almost unchanged over five successive catalytic cycles (Supplementary Figs 26–30, 31a), with the stability of the system being further verified by XRD and transmission electron microscope (TEM) measurements (Supplementary Figs 31b, 32) that show no evidence for structural or morphological differences between fresh and used catalysts. Excellent catalytic stability is also obtained with MIL-101(Fe)@Pt@MIL-101(Fe)^{9,2} (Supplementary Fig. 33).

To further confirm that the integration of Pt nanoparticles and CUSs into a single sandwich nanostructure results in useful selective hydrogenation capabilities, we used MIL-101@Pt@MIL-101 to hydrogenate smaller-sized α,β -unsaturated aldehydes. Acrolein (0.69 nm × 0.51 nm), with no substituents on C=C bond, branched 3-methyl-2-butenal (0.79 nm × 0.60 nm), and furfural (0.81 nm × 0.64 nm), with a furan ring (see Supplementary Fig. 34 for

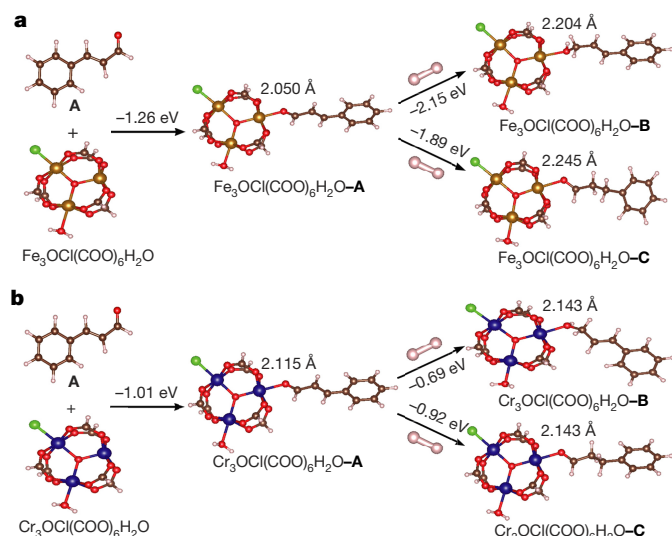


Figure 2 | Investigation of the reaction mechanism on MIL-101@Pt. a, b, Theoretical calculations of the interaction of cinnamaldehyde (A) with Fe trimers (a) or Cr trimers (b), and the subsequent selective hydrogenation of A to cinnamyl alcohol (B) and hydrocinnamaldehyde (C), on a MIL-101(Fe)@Pt or MIL-101(Cr)@Pt catalyst. Brown, carbon; pink, hydrogen; red, oxygen; gold, iron; blue, chromium; green, chlorine.

structures), are all converted through preferential hydrogenation of the C=O group over the C=C group (Table 1, Supplementary Tables 5–7 and Supplementary Figs 35–37). Conversion efficiencies are 52.7%, 59.9% and 98.5%, respectively, with selectivities for hydrogenation of the C=O group being 97.3%, 92.5% and 99.8%. When compared with other catalysts^{26,27}, MIL-101(Fe)@Pt/MIL-101(Fe)^{22,0} also exhibits excellent selectivity towards allyl alcohol (Table 1, entries 24 and 25; Supplementary Note 5 and Supplementary Table 8).

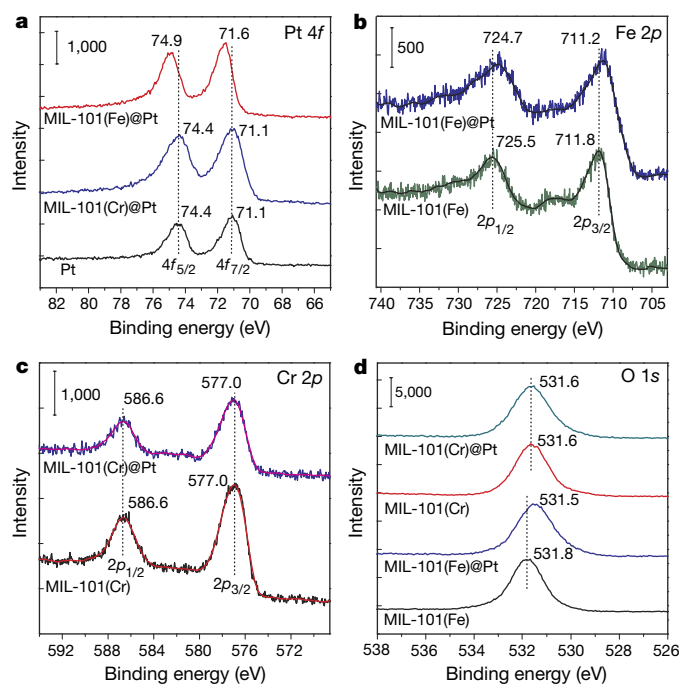


Figure 3 | XPS profiles of Pt, Fe, Cr and O in different catalysts. a, Pt 4f level. b, Fe 2p level. c, Cr 2p level. d, O 1s level. The catalysts investigated were pure Pt NPs, MIL-101(Fe), MIL-101(Cr), MIL-101(Cr)@Pt and MIL-101(Fe)@Pt. Numbers adjacent to dashed lines represent the peaks' binding energy values.

To explore the generality of the catalyst preparation method, we also synthesized^{28–30} the sandwich structures MOF-525(Zr)@Pt@MOF-525(Zr)^{26,5}, MOF-74(Co)@Pt@MOF-74(Co)^{8,4}, UiO-66(Zr)@Pt@UiO-66(Zr)^{11,2}, and UiO-67(Zr)@Pt@UiO-67(Zr)^{24,1} (Supplementary Table 2 and Supplementary Figs 38–45), which catalysed the hydrogenation of A with conversion efficiencies of 35.2%, 35.0%, 90.0% and 69.4%, respectively, and selectivities towards B of 85.0%, 70.1%, 65.0% and 73.0%, respectively (Supplementary Figs 46–49 and Supplementary Table 9). Finally, as an example of a system containing different metal nanoparticles, we also prepared MIL-101(Fe)@Ru@MIL-101(Fe)^{8,3}, which hydrogenated A with a conversion efficiency of 48.7% and a selectivity towards B of 58.5% (Supplementary Figs 50–52 and Supplementary Table 10). These results illustrate the generality of the concept of using CUSs in MOFs as a means of tuning the selectivity of the hydrogenation of α,β -unsaturated aldehydes. Moreover, using commercial Pt/C (carbon) and Pt/Fe₂O₃ for the hydrogenation of A (Supplementary Figs 53, 54) resulted in conversion efficiencies of 98.1% and 54.5% and selectivities towards B of 39.9% and 84.5%, respectively (Supplementary Table 11), with the different selectivities suggesting that Fe-based supports favour the hydrogenation of C=O groups over C=C groups (although not to the extent seen with the sandwich forms of catalysts). Taken together, our observations confirm the considerable potential of MOFs as a new generation of heterogeneous catalytic supports that may prove effective when targeting important but highly challenging reactions.

Received 14 June 2015; accepted 10 August 2016.

Published online 5 October; corrected online 2 November 2016

(see full-text HTML version for details).

- Gallezot, P. & Richard, D. Selective hydrogenation of α,β -unsaturated aldehydes. *Catal. Rev. Sci. Eng.* **40**, 81–126 (1998).
- Tian, Z., Xiang, X., Xie, L. & Li, F. Liquid-phase hydrogenation of cinnamaldehyde: enhancing selectivity of supported gold catalysts by incorporation of cerium into the support. *Ind. Eng. Chem. Res.* **52**, 288–296 (2013).
- Kahsar, K. R., Schwartz, D. K. & Medlin, J. W. Control of metal catalyst selectivity through specific noncovalent molecular interactions. *J. Am. Chem. Soc.* **136**, 520–526 (2014).
- Wu, B., Huang, H., Yang, J., Zheng, N. & Fu, G. Selective hydrogenation of α,β -unsaturated aldehydes catalyzed by amine-capped platinum-cobalt nanocrystals. *Angew. Chem. Int. Edn* **51**, 3440–3443 (2012).
- Stassi, J. P., Zgolicz, P. D., de Miguel, S. R. & Scelzo, O. A. Formation of different promoted metallic phases in PtFe and PtSn catalysts supported on carbonaceous materials used for selective hydrogenation. *J. Catal.* **306**, 11–29 (2013).
- Kennedy, G., Baker, L. R. & Somorjai, G. A. Selective amplification of C=O bond hydrogenation on Pt/TiO₂: catalytic reaction and sum-frequency generation vibrational spectroscopy studies of crotonaldehyde hydrogenation. *Angew. Chem. Int. Edn* **53**, 3405–3408 (2014).
- Kliwer, C. J., Bieri, M. & Somorjai, G. A. Hydrogenation of the α,β -unsaturated aldehydes acrolein, crotonaldehyde, and prenal over Pt single crystals: a kinetic and sum-frequency generation vibrational spectroscopy study. *J. Am. Chem. Soc.* **131**, 9958–9966 (2009).
- Ide, M. S., Hao, B., Neurock, M. & Davis, R. J. Mechanistic insights on the hydrogenation of α,β -unsaturated ketones and aldehydes to unsaturated alcohols over metal catalysts. *ACS Catal.* **2**, 671–683 (2012).
- Cargnello, M. *et al.* Control of metal nanocrystal size reveals metal-support interface role for ceria catalysts. *Science* **341**, 771–773 (2013).
- Corra, A. & Serna, P. Chemoselective hydrogenation of nitro compounds with supported gold catalysts. *Science* **313**, 332–334 (2006).
- Ma, C. Y. *et al.* Mesoporous Co₃O₄ and Au/Co₃O₄ catalysts for low-temperature oxidation of trace ethylene. *J. Am. Chem. Soc.* **132**, 2608–2613 (2010).
- Fu, Q. *et al.* Interface-confined ferrous centers for catalytic oxidation. *Science* **328**, 1141–1144 (2010).
- Vayssilov, G. N. *et al.* Support nanostructure boosts oxygen transfer to catalytically active platinum nanoparticles. *Nature Mater.* **10**, 310–315 (2011).
- Ghosh, S. *et al.* Selective oxidation of propylene to propylene oxide over silver-supported tungsten oxide nanostructure with molecular oxygen. *ACS Catal.* **4**, 2169–2174 (2014).
- Furukawa, H., Cordova, K. E., O'Keeffe, M. & Yaghi, O. M. The chemistry and applications of metal-organic frameworks. *Science* **341**, 1230444 (2013).
- McDonald, T. M. *et al.* Cooperative insertion of CO₂ in diamine-appended metal-organic frameworks. *Nature* **519**, 303–308 (2015).
- Dhakshinamoorthy, A. & Garcia, H. Catalysis by metal nanoparticles embedded on metal-organic frameworks. *Chem. Soc. Rev.* **41**, 5262–5284 (2012).

18. Lu, G. *et al.* Imparting functionality to a metal-organic framework material by controlled nanoparticle encapsulation. *Nature Chem.* **4**, 310–316 (2012).
19. Férey, G. *et al.* A chromium terephthalate-based solid with unusually large pore volumes and surface area. *Science* **309**, 2040–2042 (2005).
20. Taylor-Pashow, K. M. L., Rocca, J. D., Xie, Z., Tran, S. & Lin, W. Postsynthetic modifications of iron-carboxylate nanoscale metal-organic frameworks for imaging and drug delivery. *J. Am. Chem. Soc.* **131**, 14261–14263 (2009).
21. Bauer, S. *et al.* High-throughput assisted rationalization of the formation of metal organic frameworks in the iron (III) aminoterephthalate solvothermal system. *Inorg. Chem.* **47**, 7568–7576 (2008).
22. Xiao, D. J. *et al.* Oxidation of ethane to ethanol by N₂O in a metal-organic framework with coordinatively unsaturated iron (II) sites. *Nature Chem.* **6**, 590–595 (2014).
23. Hu, P., Morabito, J. V. & Tsung, C.-K. Core-shell catalysts of metal nanoparticle core and metal-organic framework shell. *ACS Catal.* **4**, 4409–4419 (2014).
24. Guo, Z. *et al.* Carbon nanotube-supported Pt-based bimetallic catalysts prepared by a microwave-assisted polyol reduction method and their catalytic applications in the selective hydrogenation. *J. Catal.* **276**, 314–326 (2010).
25. Nørskov, J. K. *et al.* Trends in the exchange current for hydrogen evolution. *J. Electrochem. Soc.* **152**, J23–J26 (2005).
26. Zhu, Y., Qian, H., Drake, B. A. & Jin, R. Atomically precise Au₂₅(SR)₁₈ nanoparticles as catalysts for the selective hydrogenation of α,β -unsaturated ketones and aldehydes. *Angew. Chem. Int. Edn* **49**, 1295–1298 (2010).
27. Cano, I., Chapman, A. M., Urakawa, A. & van Leeuwen, P. W. N. M. Air-stable gold nanoparticles ligated by secondary phosphine oxides for the chemoselective hydrogenation of aldehydes: crucial role of the ligand. *J. Am. Chem. Soc.* **136**, 2520–2528 (2014).
28. Morris, W. *et al.* Synthesis, structure, and metalation of two new highly porous zirconium metal-organic frameworks. *Inorg. Chem.* **51**, 6443–6445 (2012).
29. Glover, T. G., Peterson, G. W., Schindler, B. J., Britt, D. & Yaghi, O. M. MOF-74 building unit has a direct impact on toxic gas adsorption. *Chem. Eng. Sci.* **66**, 163–170 (2011).
30. Na, K., Choi, K. M., Yaghi, O. M. & Somorjai, G. A. Metal nanocrystals embedded in single nanocrystals of MOFs give unusual selectivity as heterogeneous catalysts. *Nano Lett.* **14**, 5979–5983 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported financially by the National Research Fund for Fundamental Key Project (grant 2014CB931801 to Z.T.), the National Natural Science Foundation of China (grants 21475029 and 91427302 to Z.T., and 21303029 to G.L.), the Instrument Developing Project of the Chinese Academy of Sciences (grant YZ201311 to Z.T.), the CAS-CSIRO Cooperative Research Program (grant GJHZ1503 to Z.T.), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDA09040100 to Z.T.), and the Youth Innovation Promotion Association CAS (grant 2016036 to G.L.). We gratefully acknowledge the use of the supercomputer facilities at the National Computational Infrastructure (NCI) in Canberra, Australia.

Author Contributions Z.T. and G.L. proposed the research direction and guided the project. M.Z., K.Y. and G.L. designed and performed the materials synthesis, characterization, and catalytic tests. Y.W. and H.Z. performed the theoretical calculations and explained the catalytic results. J.G. helped to detect and analyse the Brunauer–Emmett–Teller (BET) surface areas and pore-size distributions of samples. L.G. guided the high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) imaging of sandwich structures. W.H. took part in the characterization of some samples, and discussed the results. Z.T., G.L., M.Z. and Y.W. drafted the manuscript. All of the authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Z.T. (zytang@nanoctr.cn), G.L. (liguodong@nanoctr.cn) or H.Z. (h.zhao@griffith.edu.au).

Reviewer Information *Nature* thanks P. Claus and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Mantle dynamics inferred from the crystallographic preferred orientation of bridgmanite

Noriyoshi Tsujino¹, Yu Nishihara², Daisuke Yamazaki¹, Yusuke Seto³, Yuji Higo⁴ & Eiichi Takahashi⁵

Seismic shear wave anisotropy^{1–6} is observed in Earth's uppermost lower mantle around several subducted slabs. The anisotropy caused by the deformation-induced crystallographic preferred orientation (CPO) of bridgmanite (perovskite-structured (Mg,Fe)SiO₃) is the most plausible explanation for these seismic observations. However, the rheological properties of bridgmanite are largely unknown. Uniaxial deformation experiments^{7–9} have been carried out to determine the deformation texture of bridgmanite, but the dominant slip system (the slip direction and plane) has not been determined. Here we report the CPO pattern and dominant slip system of bridgmanite under conditions that correspond to the uppermost lower mantle (25 gigapascals and 1,873 kelvin) obtained through simple shear deformation experiments using the Kawai-type deformation-DIA apparatus¹⁰. The fabrics obtained are characterized by [100] perpendicular to the shear plane and [001] parallel to the shear direction, implying that the dominant slip system of bridgmanite is [001](100). The observed seismic shear-wave anisotropies near several subducted slabs^{1–4} (Tonga–Kermadec, Kurile, Peru and Java) can be explained in terms of the CPO of bridgmanite as induced by mantle flow parallel to the direction of subduction.

Recent P-wave seismic tomography¹¹ revealed four geometries of subducted slabs in the mantle. Type I refers to a slab that is stagnant above the 660-km discontinuity. Type II is assigned to slabs that penetrate the 660-km discontinuity (for example, Kuril). Type III slabs lie in the uppermost lower mantle at depths between 660–1,000 km (for example, Tonga–Kermadec, Sumatra and South America). Type IV slabs continuously descend into the mid lower mantle. Seismic tomography^{11–13} provides morphological images of the subducted slabs in Earth's mantle, but does not give any insight into the dynamic features of the slab. However, the seismic anisotropy observed in the mantle provides information on the flow direction of the slabs in the deep mantle because this anisotropy may be the reflection of the CPO of the constituent minerals, which is yielded by dislocation creep deformation.

As shown in previous reports, an important anisotropy that demonstrates clear shear-wave splitting was observed in the uppermost lower mantle of the Tonga–Kermadec slab region^{1,2}. Recently, such anisotropies were also observed in the uppermost lower mantle of other slab regions^{3–5}. In these observations, shear wave splitting shows fast polarization that is generally parallel to the plane of the subducted plate (for example, Tonga–Kermadec, Java, Peru and Kurile)^{1–4}, whereas other observed slab regions do not show a consistent splitting pattern³. It is therefore interesting to examine how the observed shear-wave anisotropy could be interpreted by the CPO of the lower-mantle minerals. It is generally accepted that in the pyrolitic mantle model, the constituents of the lower mantle are bridgmanite or (Mg,Fe)SiO₃ perovskite, of space group *Pbnm* (77 vol%), ferropericlase (16 vol%) and CaSiO₃ perovskite (7 vol%)¹⁴. The contribution of the CPO of ferropericlase to seismic anisotropy is negligible because of its nearly isotropic elasticity

at the uppermost lower mantle conditions¹⁵. CaSiO₃ perovskite is elastically anisotropic¹⁶ but the contribution of this phase to seismic anisotropy will not be substantial because of the small amount of CaSiO₃ in the lower mantle. In contrast, it is anticipated that considerable seismic anisotropy can be produced by the CPO of bridgmanite^{17,18} on the basis of its high elastic anisotropy^{19,20} and overwhelming presence in the lower mantle. We therefore focused on the deformation-induced CPO of bridgmanite as the most important factor in lower-mantle seismic anisotropy.

We performed shear deformation experiments on bridgmanite with a controlled strain rate under lower-mantle conditions by employing the deformation-DIA-type multi-anvil press with the Kawai-type cell assembly (6–8 type)¹⁰. Dense, sintered (Mg_{0.97}Fe_{0.03})SiO₃ bridgmanite aggregates synthesized at 25 GPa and 1,873 K were used as samples for the deformation experiments. A backscattered electron image (Fig. 1a) and pole figures (Fig. 2a) of sintered bridgmanite revealed that the sample was an equigranular aggregate with a typical grain size of 15 μm and a random crystallographic orientation, determined from two-dimensional monochromatic X-ray diffraction patterns (see Methods). In the deformation experiments, a thin ellipse of the bridgmanite aggregate with a strain marker of Ni foil (which was initially set at the middle of the sample, perpendicular to the cut surfaces of the pistons) was sandwiched between 45°-cut polycrystalline dense alumina pistons at the centre of the cell assembly. As the differential rams of the deformation-DIA press advance, the alumina pistons apply simple shear to the thin bridgmanite sample.

Figure 1b and c shows the microstructures of the recovered samples from the undeformed (annealing only) experiment (run K116) and from the deformation experiment (run K122), respectively, both performed at identical pressure (*P*) and temperature (*T*) conditions (25 GPa, 1,873 K). In the former, the strain marker was almost normal to the cut surfaces of the alumina pistons, showing that the shear strain of the bridgmanite sample was very close to zero. The sample was thus almost free from deformation throughout the experimental processes; that is, compression, heating, annealing and subsequent decompression. As shown in Fig. 1c, the strain marker definitely tilted in the deformed sample. From the tilting angle of the strain marker (~38°) and the deformation time (1 h), the total strain and average strain rate were determined to be $\gamma \approx 0.8$ and $\dot{\gamma} \approx 2 \times 10^{-4} \text{ s}^{-1}$, respectively. The thickness of the deformed sample was reduced by about 20% in comparison with the original, suggesting that a uniaxial component of deformation was present. The microstructural observation in Fig. 1d shows that elongated bridgmanite grains of sizes up to 30–40 μm accumulated in the shear direction and are surrounded by small grains of less than a few micrometres. The structural features indicate that dynamic recrystallization occurred. The CPO of bridgmanite could therefore be induced by deformation involving dislocations.

The crystallographic orientation of the bridgmanite was determined using the two-dimensional monochromatic X-ray diffraction pattern

¹Institute for Planetary Materials, Okayama University, 827 Yamada, Misasa, Tottori 682-0193, Japan. ²Geodynamics Research Center, Ehime-University, 2-5 Bunkyo-cho, Matsuyama, Ehime, 790-8577, Japan. ³Department of Planetology, Kobe University, Kobe 657-8501, Japan. ⁴Japan Synchrotron Radiation Research Institute, 1-1-1 Kouto, Sayo, Hyogo 689-5198, Japan.

⁵Department of Earth and Planetary Sciences, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo 152-8551, Japan.

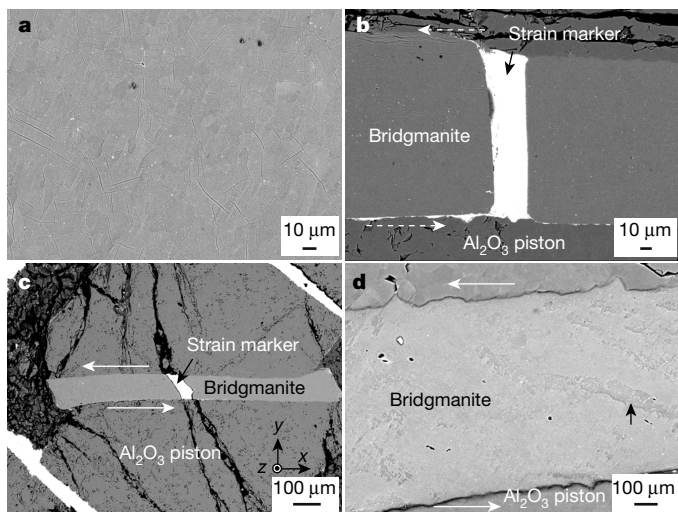


Figure 1 | Backscattered electron images of the bridgmanite aggregates. **a**, Sintered sample. **b**, Sample annealed at 25 GPa and 1,873 K for 10 min (run K116). **c**, **d**, Sample deformed at 25 GPa and 1,873 K for 1 h (run K122) before (c) and after (d) etching with colloidal silica. In **a** and **d** the polished surfaces were etched with colloidal silica to clarify the grain boundaries. In **b** and **c** the black arrows indicate the Ni foil strain markers. The strain marker in the annealed sample in **b** remains perpendicular to the 45°-cut surfaces of the alumina pistons, indicating almost no deformation during compression and annealing. In **c**, the Ni strain marker in the deformed sample is tilted by 38°, indicating a strain of $\gamma \approx 0.8$. Elongated grains, indicating dynamic recrystallization, were observed only in the deformed sample (black arrow in **d**). The x , y and z axes correspond to the shear direction, the shear plane normal and direction perpendicular to both x and y , respectively. This definition of coordinates is used in Figs 2 and 3 and in Extended Data Fig. 7.

method^{7,9,21} for both the starting and deformed samples. The sheared bridgmanite aggregate demonstrated strong fabric (Fig. 2b), whereas the starting material did not show a notable CPO pattern (Fig. 2a). In the sheared bridgmanite, the [100] axis was oriented perpendicular to the shear plane. Although weak girdle patterns of the [001] and [010] axes (which correspond to uniaxial compressive strain) were observed, the [001] and [010] axes were mostly aligned parallel and normal to the shear direction on the shear plane, respectively. Experimental textures are compatible with orientation-producing deformation mechanisms such as dominant slip on [001](100) at 25 GPa and 1,873 K, which correspond to conditions found in the uppermost lower mantle.

The dominant slip system of bridgmanite at high deviatoric stress has been reported through high-pressure experiments^{7,8,22} and theoretical calculations^{18,23}, as shown in Extended Data Table 1. Using a diamond anvil cell, uniaxial deformation experiments were conducted on bridgmanite at room temperature^{7,22}. No clear CPO pattern was observed⁷, probably owing to insufficient strain, whereas dominant slip systems were reported as [100], [010] and $\langle 110 \rangle$ on the (001) plane below 55 GPa and the (100) plane over 55 GPa (ref. 22). Uniaxial stress relaxation experiments (USRE) on bridgmanite were carried out using the Kawai-type multi-anvil press at 25 GPa and 1,673 K (ref. 8). Analysis of the recovered samples using the X-ray peak broadening technique suggested that the dominant slip direction is [100], in contradiction with our results. Dense deformation bands across the pre-existing twin boundaries were observed in the USRE⁸, suggesting that the bridgmanite aggregate was deformed under very high deviatoric stress. Extremely high deviatoric stress values of up to several gigapascals, induced by compression at room temperature, were directly measured during *in situ* USRE of olivine polymorphs^{24,25}. The [100] slip direction observed in USRE of bridgmanite⁸ might be the result of very high deviatoric stress.

According to first-principles calculations²³ the [100](001) slip system is one of the most easily activated slip systems of bridgmanite in the dislocation glide region. This supports the hypothesis that the slip direction observed in the USRE results from dislocation glide. On the other hand, another study based on first-principles calculations with a visco-plastic self-consistent model¹⁸ suggested that the most easily activated slip system is [010](100) at 0 K for a pressure range of 0–100 GPa in the dislocation glide region. Both calculations and previous experimental results are inconsistent with the present results. This could be due to differences in the dominant slip mechanisms or the dominant deformation mechanisms found for varying experimental conditions (for example, stress, temperature and pressure), such as in the case of olivine²⁶.

CaTiO₃ perovskite is often used as an analogue material for bridgmanite because the SiO₆ or TiO₆ octahedra have similar rotation angles. Uniaxial deformation experiments were performed on CaTiO₃ perovskite²⁷, in which dislocation creep was expected owing to the high temperature (up to 1,973 K) and low deviatoric stress (25–120 MPa). TEM observations of the recovered sample revealed that screw dislocation with Burgers vectors [100]_{pc} and [011]_{pc} on the (01–1)_{pc} plane, where indexing is based on the pseudo-cubic system, formed rectangular networks. The [100]_{pc} and [011]_{pc} directions on the (01–1)_{pc} plane in pseudo-cubic indexation of perovskite include the 1/2[001] and [010] directions on the (100) plane in the orthorhombic system. The [001](100) slip system in the present study does not contradict that of CaTiO₃ perovskite.

The seismic wave anisotropy formed by the CPO of deformed bridgmanite in the present study, as shown in Fig. 3 and Extended Data Table 2, was calculated on the basis of the elastic constant (Extended Data Table 3)¹⁹. For the shear-wave anisotropy of deformed bridgmanite, the velocity of the horizontally polarized shear waves, V_{SH} , is around 1% higher than that of vertically polarized shear waves, V_{SV} , in horizontal flow, as shown in Fig. 3b, whereas V_{SV} is 0.03%–1.10% higher in vertical flow (Fig. 3d). As shown in Fig. 4a, shear-wave splitting with $V_{SH} > V_{SV}$ is observed in the uppermost lower mantle from the Tonga–Kermadec subduction zone to the Australian

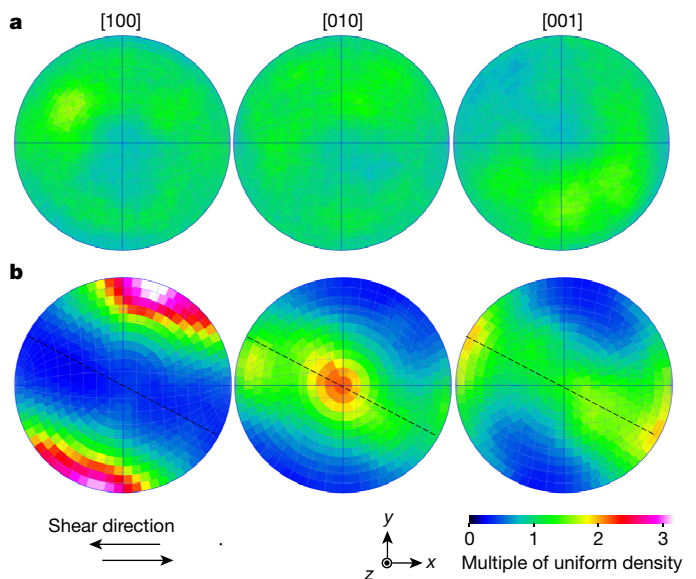


Figure 2 | Pole figures of bridgmanite showing the variation in the crystallographic orientation of the [100], [010] and [001] directions. **a**, **b**, The coordinate system is defined with respect to the deformation geometry of sintered bridgmanite (**a**) and deformed bridgmanite (K122; **b**) at $\gamma \approx 0.8$ and $\dot{\gamma} \approx 2 \times 10^{-4} \text{ s}^{-1}$. Black dashed lines represent the long axis of the ellipsoid strain of the deformed bridgmanite. The colour scale shows the intensity of the concentration of each axis as a multiple of the uniform density.

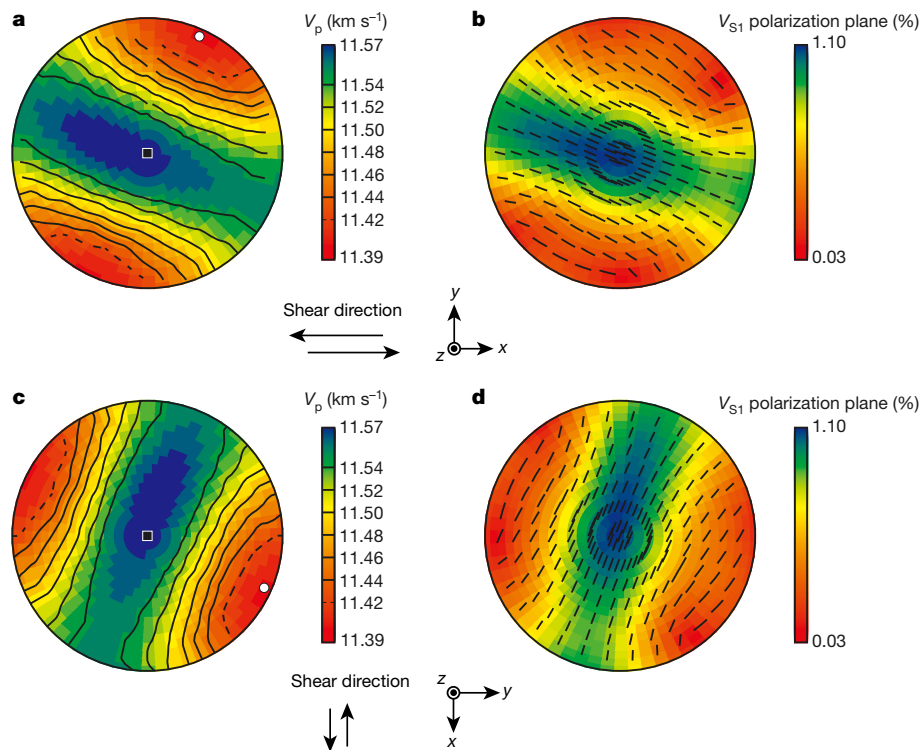


Figure 3 | Seismic wave anisotropy of bridgmanite aggregates deformed under shear at uppermost lower-mantle pressures and temperatures. These values were calculated using the ANISch5 and VpG software³². **a, c**, P-wave seismic velocity (V_p) for horizontal (**a**) and vertical (**c**) flow, respectively. **b, d**, The fastest S-wave seismic velocity (V_{s1}) polarization plane for horizontal (**b**) and vertical (**d**) flow, respectively. Black dashes show the polarization direction of the fastest S-wave.

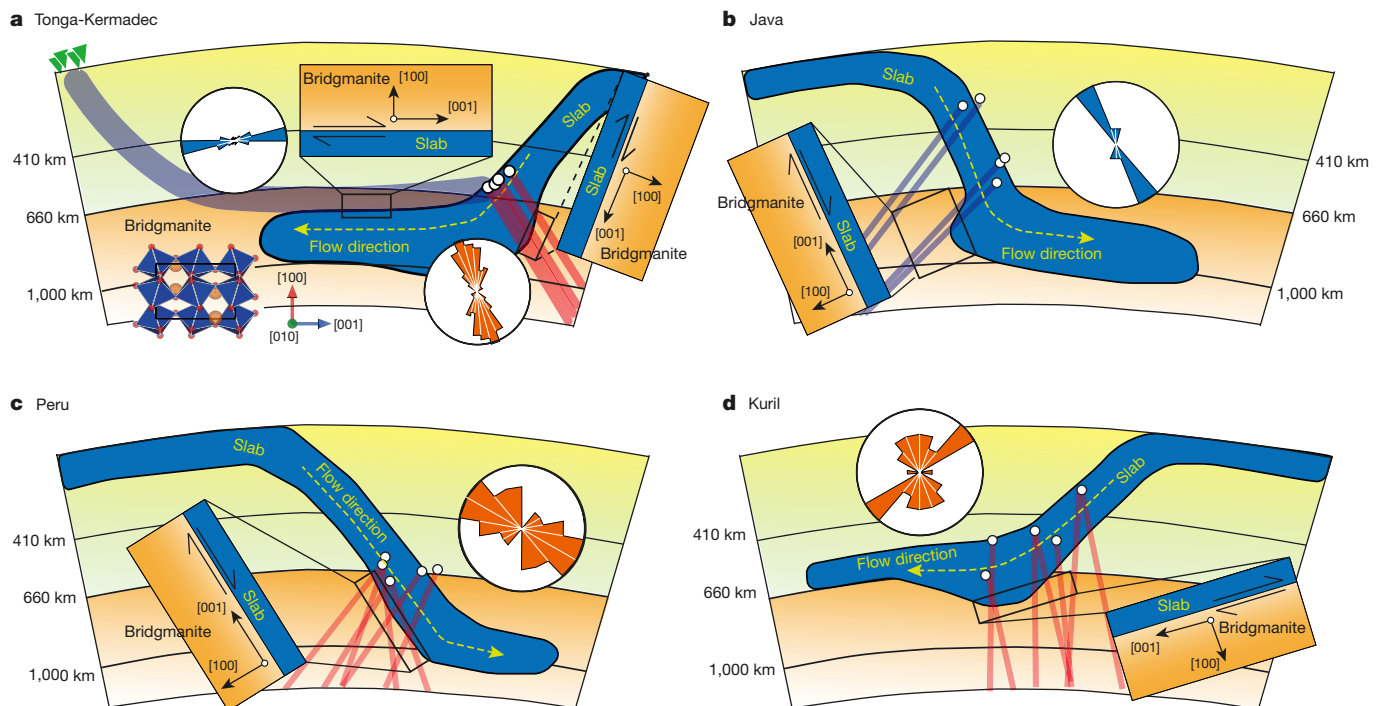


Figure 4 | Schematic cross-sections of subducted slabs. **a**, The Tonga-Kermadec arc. **b**, The Java arc. **c**, The Peru arc. **d**, The Kuril arc. All cross-sections are based on P-wave tomography¹¹. The coloured lines represent the shear seismic ray paths that are nearly parallel (blue) and perpendicular (orange) to the section for the observation of shear-wave anisotropies at the Tonga-Kermadec slab^{1,2,4} and the Kuril, Peru and Java slabs³, where the observation area, direction and stagnation depth were different. The blue and orange polar circular histograms show the polarization of the fast shear waves perpendicular to the blue and orange seismic ray paths, respectively. The fast polarization arrangement is

almost parallel to the plane of the subducted plates. Yellow dashed lines correspond to the flow directions of the subducted slab. Black double arrows represent shear directions. The uppermost lower mantle near the subducted slab is deformed by the flow directions that are parallel to the subducted slab, which align the [100] and [001] axes of bridgmanite normal to the shear deformation plane and parallel to the shear direction, respectively. The lower inset in **a** shows the crystal structure of bridgmanite. Blue octahedra are the Si sites and the orange and red spheres are the Mg and Si atoms, respectively.

continental seismic stations^{1,2}, where the ray path is nearly parallel to the section. The delay time ranges from 0.7 s to 6.2 s (refs 1 and 2), which corresponds to an average value of $(V_{SH} - V_{SV})/V_{SV} \approx 0.1\% - 1\%$, assuming a 3,000–6,000-km ray path including a 1,000–2,000-km subducted plate region¹¹ on the uppermost lower mantle. The magnitude of the shear-wave anisotropy calculated here is consistent with deformed bridgmanite (Fig. 3b). In contrast, the opposite shear wave anisotropy ($V_{SV} > V_{SH}$) was observed from the Tonga–Kermadec subduction zone at the western North America stations⁴, where the ray path is nearly perpendicular to the section. It is thus concluded that both of the observed shear-wave anisotropies around the Tonga–Kermadec subduction zone, which are type III in the P wave tomography¹¹, are well explained by the CPO of bridgmanite yielded by the penetration of subducted slabs down to 1,000 km by vertical motion and subsequent stagnation at similar depths, with a horizontal flow in the pyrolytic mantle (as shown in Fig. 4a). This flow pattern could be caused by the viscosity hill^{28–30} at depths between 1,000 km and 1,500 km, and the viscosity reduction of the slab by interconnection of ferropericlase in the post-spinel phase³¹. Shear wave anisotropies at the top of the lower mantle beneath the Java (type III), Peru (type III) and Kuril (type II) subducted slabs were observed along the seismic ray path nearly perpendicular or parallel to the subducted plate surface³ (Fig. 4b–d). In these observations, the fast polarizations in the shear-wave splitting appear to be arranged roughly parallel to the plane of the plate³, which is explained by the deformation-induced CPO of bridgmanite owing to flow parallel to the direction of subduction. The findings presented here on the CPO of bridgmanite, coupled with the observations of seismic anisotropy, provide a greater understanding of the direction of mantle flow, confirming inferences from seismic tomography data.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June 2015; accepted 15 August 2016.

Published online 17 October 2016.

1. Wookey, J., Kendall, J.-M. & Barruol, G. Mid-mantle deformation inferred from seismic anisotropy. *Nature* **415**, 777–780 (2002).
2. Wookey, J. & Kendall, J.-M. Evidence of midmantle anisotropy from shear wave splitting and the influence of shear-coupled P wave. *J. Geophys. Res.* **109**, B07309 (2004).
3. Nowacki, A., Kendall, J. M., Wookey, J. & Pemberton, A. Mid-mantle anisotropy in subduction zones and deep water transport. *Geochem. Geophys. Geosys.* **16**, 764–784 (2015).
4. Foley, B. J. & Long, M. D. Upper and mid-mantle anisotropy beneath the Tonga slab. *Geophys. Res. Lett.* **38**, L02303 (2011).
5. Lynner, C. & Long, M. D. Sub-slab anisotropy beneath the Sumatra and circum-Pacific subduction zones from source-side shear wave splitting observations. *Geochem. Geophys. Geosys.* **15**, 2262–2281 (2014).
6. Heintz, M. Midmantle deformation between the Australian continent and the Fiji–Tonga subduction zone? *J. Geophys. Res.* **111**, B09303 (2006).
7. Merkel, S. *et al.* Deformation of $(\text{Mg}_{0.9}\text{Fe}_{0.1})\text{SiO}_3$ perovskite aggregates up to 32 GPa. *Earth Planet. Sci. Lett.* **209**, 351–360 (2003).
8. Cordier, P., Ungár, T., Zsoldos, L. & Tichy, G. Dislocation creep in MgSiO_3 perovskite at conditions of the Earth's uppermost lower mantle. *Nature* **428**, 837–840 (2004).
9. Wenk, H.-R. *et al.* Deformation textures produced in diamond anvil experiments, analysed in radial diffraction geometry. *J. Phys. Condens. Matter* **18**, S933–S947 (2006).
10. Tsujino, N. *Experimental Study on Rheology of the Earth's Lower Mantle*. PhD thesis, Tokyo Institute of Technology, (2012).
11. Fukao, Y. & Obayashi, M. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity. *J. Geophys. Res.* **118**, 5920–5938 (2013).
12. van der Hilst, R. D., Widiyantoro, R. D. S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).

13. Zhao, D. Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics. *Phys. Earth Planet. Inter.* **146**, 3–34 (2004).
14. Hirose, K. Phase transitions in pyrolytic mantle around 670-km depth: implications for upwelling of plumes from the lower mantle. *J. Geophys. Res.* **107**(B4), 2078 (2002).
15. Yamazaki, D. & Karato, S. Fabric development in $(\text{Mg,Fe})\text{O}$ during large strain, shear deformation: implications for seismic anisotropy in Earth's lower mantle. *Phys. Earth Planet. Inter.* **131**, 251–267 (2002).
16. Karki, B. B. & Crain, J. First-principles determination of elastic properties of CaSiO_3 perovskite at lower mantle pressures. *Geophys. Res. Lett.* **25**, 2741–2744 (1998).
17. Wenk, H.-R., Speziale, S., MacNamara, A. K. & Garnero, E. J. Modeling lower mantle anisotropy development in a subducting slab. *Earth Planet. Sci. Lett.* **245**, 302–314 (2006).
18. Mainprice, D., Tommasi, A., Ferré, D., Carres, P. & Cordier, P. Predicted glide systems and crystal preferred orientations of polycrystalline silicate Mg-Perovskite at high pressure: implications for the seismic anisotropy in the lower mantle. *Earth Planet. Sci. Lett.* **271**, 135–144 (2008).
19. Wentzcovitch, R. M., Karki, B. B., Cococcioni, M. & de Gironcoli, S. Thermoelastic properties of MgSiO_3 -perovskite: insights on the nature of the Earth's lower mantle. *Phys. Rev. Lett.* **92**, 018501 (2004).
20. Oganov, A. R., Brodholt, J. P. & Price, G. D. The elastic constants of MgSiO_3 perovskite at pressures and temperatures of the Earth's mantle. *Nature* **411**, 934–937 (2001).
21. Seto, Y. Whole pattern fitting for two-dimensional diffraction patterns from polycrystalline materials. *Rev. High Press. Sci. Technol.* **22**, 144–152 (2012).
22. Miyagi, L. & Wenk, H.-R. Texture development and slip systems in bridgmanite and bridgmanite + ferropericlase aggregates. *Phys. Chem. Miner.* **43**, 597–613 (2016).
23. Ferré, D., Carrez, P. & Cordier, P. First principles determination of dislocations properties of MgSiO_3 perovskite at 30 GPa based on the Peierls–Nabarro model. *Phys. Earth Planet. Inter.* **163**, 283–291 (2007).
24. Chen, J., Inoue, T., Weidner, D. J., Wu, Y. & Vaughan, M. T. Strength and water weakening of mantle minerals olivine, wadsleyite and ringwoodite. *Geophys. Res. Lett.* **25**, 575–578 (1998).
25. Nishihara, Y. *et al.* Stress relaxation experiments of olivine under conditions of subducted slab in Earth's deep upper mantle. *Phys. Earth Planet. Inter.* **183**, 164–174 (2010).
26. Jung, H. & Karato, S. Water-induced fabric transitions in olivine. *Science* **293**, 1460–1463 (2001).
27. Besson, P., Poirier, J. P. & Price, G. D. Dislocations in CaTiO_3 perovskite deformed at high-temperature: a transmission electron microscopy study. *Phys. Chem. Miner.* **23**, 337–344 (1996).
28. Forte, M. A. & Mitrovica, J. X. Deep-mantle high-viscosity flow and thermochemical structure inferred from seismic and geodynamic data. *Nature* **410**, 1049–1056 (2001).
29. Morra, G. *et al.* The fate of the slabs interacting with a density/viscosity hill in the mid-mantle. *Phys. Earth Planet. Inter.* **180**, 271–282 (2010).
30. Marquardt, H. & Miyagi, L. Slab stagnation in the shallow lower mantle linked to an increase in mantle viscosity. *Nat. Geosci.* **8**, 311–314 (2015).
31. Yamazaki, D., Yoshino, T. & Nakakuki, T. Interconnection of ferro-periclase controls subducted slab morphology at the top of the lower mantle. *Earth Planet. Sci. Lett.* **403**, 352–357 (2014).
32. Mainprice, D. A FORTRAN program to calculate seismic anisotropy from the lattice preferred orientation of minerals. *Comput. Geosci.* **16**, 385–393 (1990).

Acknowledgements We thank E. Ito and S. Karato for help preparing the manuscript and Takahashi Laboratory and HACTO group members for help collecting the two-dimensional diffraction data. We also thank T. Ohuchi and D. Mainprice for the codes used to calculate the seismic anisotropy of bridgmanite. This work was supported by JSPS KAKENHI Grant number 15J09669 to N.T. and 25247088 and 21109001 to E.T. The two-dimensional diffraction measurements for analysis of CPO were carried out on the BL04B1 beamline at SPring-8 under the approval with JASRI (proposal numbers 2012B1437, 2013B1434, 2014A1431, 2014B1400, 2015A1600 and 2015B1504).

Author Contributions N.T. planned and performed the deformation experiments (with help from Y.N., D.Y. and E.T.). N.T. and D.Y. collected two-dimensional diffraction patterns (with help from Y.H.). N.T. analysed the crystal orientation distribution of the bridgmanite aggregate (with help from Y.S.). N.T., D.Y. and Y.N. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.T. (tsujino@okayama-u.ac.jp).

METHODS

Preparation of the starting material of bridgmanite aggregates. Well-sintered ($\text{Mg}_{0.97}\text{Fe}_{0.03}$) SiO_3 -bridgmanite aggregates that were free from a CPO were prepared using the following procedures. The composition is very close to the bridgmanite in the post-spinel assembly formed from San Carlos olivine ($\text{Mg}_{0.9}\text{Fe}_{0.1}$) SiO_4 (ref. 33). First an orthopyroxene (opx) powder with the composition of ($\text{Mg}_{0.97}\text{Fe}_{0.03}$) SiO_3 was synthesized from a mixture of MgO, SiO_2 and Fe_2O_3 with the prescribed ratio at 1,673 K and a fugacity f_{O_2} of fayalite–magnetite–quartz (FMQ; 1 log unit) using a gas mixture (H_2 and CO_2) furnace. The opx powder was then put in an iron inner capsule and sintered in a piston-cylinder apparatus at 1 GPa and 1,473 K. To avoid the formation of cracks in the sintered opx, we used pyrex glass as an outer capsule and decompressed the sample slowly at temperatures higher than 1,073 K after sintering. Bridgmanite aggregates for the deformation experiments were synthesized from the sintered opx at 25 GPa and 1,873 K for 1 h in the Kawai-type apparatus. It is known that substantial amounts of CPO develop in the bridgmanite aggregate, accompanied by conversion from opx under high deviatoric stress conditions⁹. In the synthesis process used for bridgmanite, we therefore use an NaCl capsule to reduce the deviatoric stress³⁴. The obtained bridgmanite aggregates were examined by field emission scanning electron microscopy and 2D X-ray diffraction (see Sample characterization section). The backscattered electron image (Fig. 1a) and the pole figures (Fig. 2b) of the sintered bridgmanite aggregates confirmed that the aggregates had a random crystallographic orientation with a grain-size of around 15 μm and are suitable as a starting material for the shear deformation experiments and subsequent CPO analyses.

Deformation experiments. Shear deformation experiments at lower mantle conditions were conducted using the deformation-DIA type apparatus, known as the Kawai-type apparatus for triaxial deformation (KATD)³⁵, installed at the Tokyo Institute of Technology. Extended Data Fig. 1a shows a schematic of the cell assembly adopted in the present deformation experiments. A cylinder of LaCrO_3 was used as a heater. The upper and bottom pistons of well-sintered hard alumina are cut to make 45° surfaces that are used to shear stress to the sample mounted between them when the differential rams are driven. A Pt foil (50 μm thick) was placed at the ends of the 45°-cut alumina pistons to reduce the friction against the sideslip. An Ni foil (30 μm thick) set at the middle of the ellipse sample served as a strain marker, which also kept the oxygen fugacity of the sample equal to or less than the Ni–NiO buffer. The temperature was determined from the electric power on the basis of calibrations performed using a similar cell assembly with thermocouple (Extended Data Fig. 1b). Extended Data Fig. 2 shows the relationship between the generated temperature and the electric power obtained in two calibration runs (K63 and K71). It should be noted that both curves almost overlapped each other up to 1,873 K. Pressure calibration of the KATD apparatus was carried out by detecting the phase transformations from α – β in Mg_2SiO_4 (at 15.1 GPa)³⁶, β – γ in Mg_2SiO_4 (at 19.8 GPa)³⁷ and the dissociation of γ – Mg_2SiO_4 to MgSiO_3 –bridgmanite + MgO (periclase) (at 23.6 GPa)³⁸ at a temperature of 1,873 K.

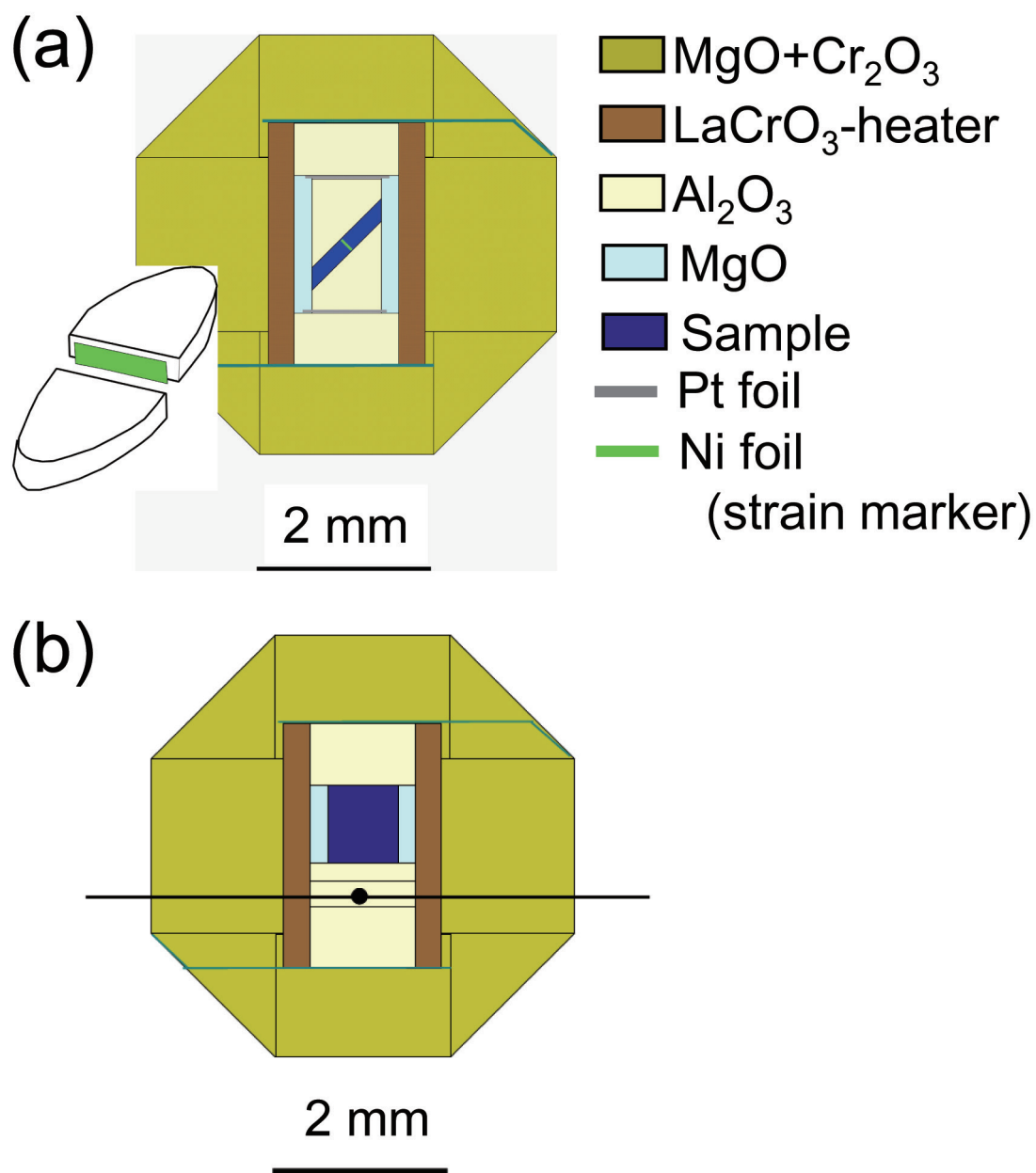
The specimens were first compressed to the desired pressure (25 GPa) at room temperature, and then heated at 1,873 K for 10 min to relax the deviatoric stress in the specimens. In run K116, which is the undeformed experiment, the specimens were quenched by shutting off the electric power and recovered to observe the state of the sample just before the deformation experiment. In the deformation experiment run (K122), the sample was also first annealed at 1,873 K for 10 min and then deformed by advancing each of the upper and lower differential rams to a displacement of 75 μm (that is, 150 μm in total) for a duration of 1 h at 1,873 K (see Extended Data Fig. 3). Throughout the heating and the deformation stages, the load of the main ram was kept constant. The loads of the differential rams were linearly increased to advance the differential rams at high temperature (1,873 K). High-pressure experiments were conducted under nominally dry conditions. As the water solubility of Al-free bridgmanite is very low³⁹ (<1 parts per million), the effect of water is considered to be negligible in the present study.

Sample characterization. To observe the microstructure, the starting material and the recovered sample were polished with SiC sand papers and diamond paste in

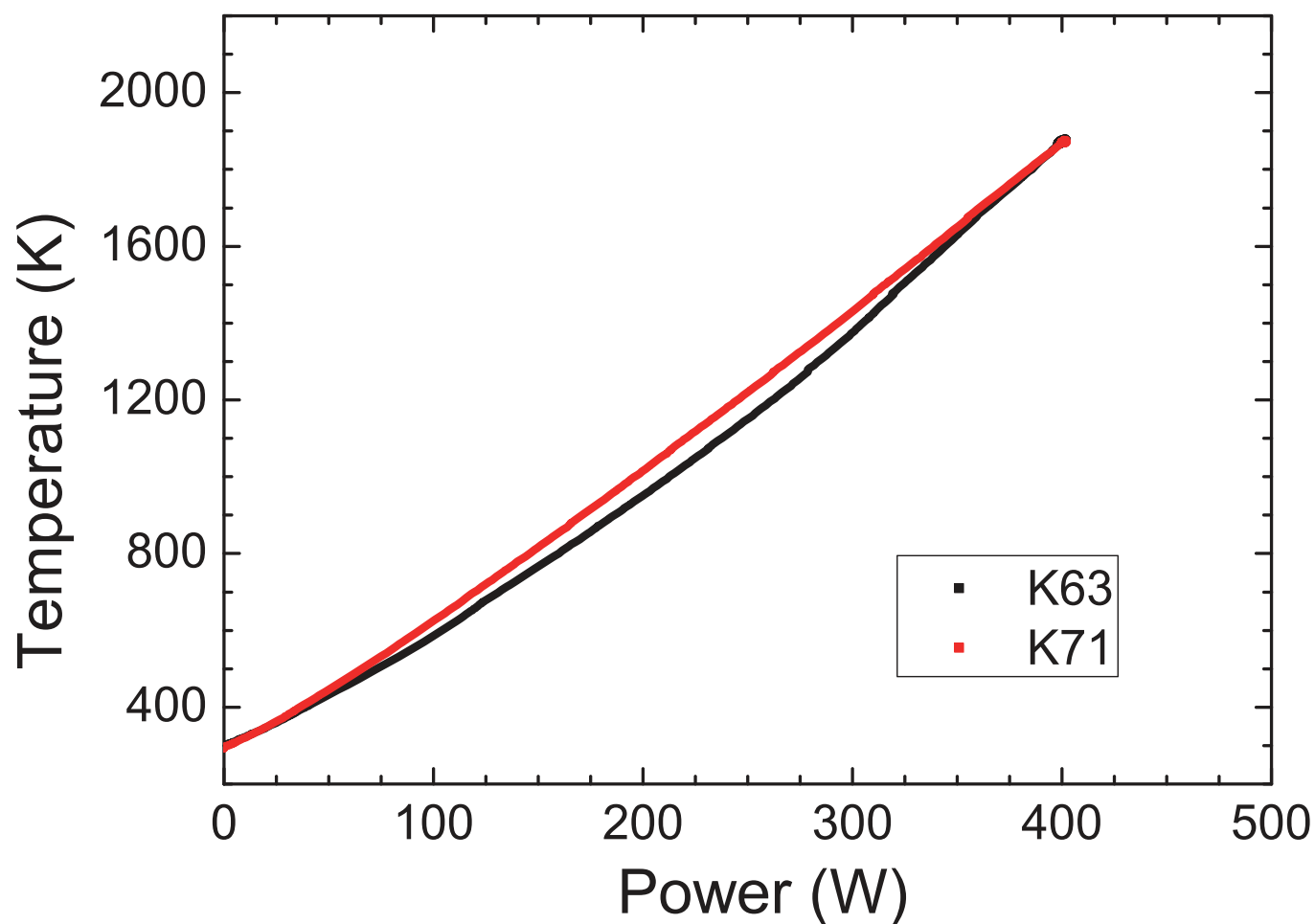
sequence, and were then etched by colloidal silica to clarify the grain boundaries. The microstructures of the samples were investigated by a field emission scanning electron microscope (JSM-7001F) at ISEI, Okayama University, Japan, as shown in Fig. 1.

The electron beam used in the electron backscattered diffraction technique that is usually applied to measure crystallographic orientation would cause serious damage to the bridgmanite crystals. Therefore, we determined the crystallographic orientation of the bridgmanite by the two-dimensional monochromatic X-ray diffraction pattern method^{7,9,21} for both the starting and the deformed samples. The two-dimensional monochromatic X-ray diffraction patterns were acquired by the imaging plate (FUJIFILM, 200 mm \times 250 mm) at the synchrotron beam line, BL04B1, of SPring-8 at the Japan Synchrotron Radiation Research Institute (JASRI), Hyogo, Japan⁴⁰. A monochromatic X-ray of 61.388 keV (wavelength $\lambda = 0.20197 \text{ \AA}$) was employed. X-ray diffraction measurements were collected using a beam size of 100 μm \times 200 μm and a sample length of 500 μm . As shown in Extended Data Fig. 4, the direction of the X-ray beam is parallel to the shear plane and perpendicular to the shear direction. The zero degree azimuth angle ϕ indicates that it is perpendicular to the shear deformation direction. The distance between the sample and the imaging plate was about 601.8 mm, calibrated using a CeO_2 standard. The typical acquisition time was 10 min. The imaging plate data were digitized using a FUJI BAS2000 reader under a resolution of 100 μm \times 100 μm . Extended Data Fig. 5 represents the two-dimensional and one-dimensional X-ray diffraction patterns converted from IP data. In the CPO analysis of the bridgmanite aggregates, the (111), (020), (120), (210), (022), (202), (113), (122), (212), (023) and (221) diffraction peaks were adopted. The software ReciPro²¹ was used for the CPO analysis (this software was previously used for CPO analysis in other studies^{41,42}). The analysis was carried out under the following parameters: number of crystallites (2,000,000); size of the crystallites (1 μm); beam convergence angle (0.02°); beam monochromaticity (0.1%); step increment (1%); and directional density (60%). As shown in Extended Data Fig. 6, the observed intensities for all peaks are in good agreement with those of the simulated peaks. In addition, the CPO results were reanalysed by the MAUD program. As shown in Extended Data Fig. 7, the CPO of the deformed bridgmanite from the ReciPro and MAUD software programs were confirmed to be same, although the intensities of the CPO are not perfectly identical because of the different algorithms used in each program.

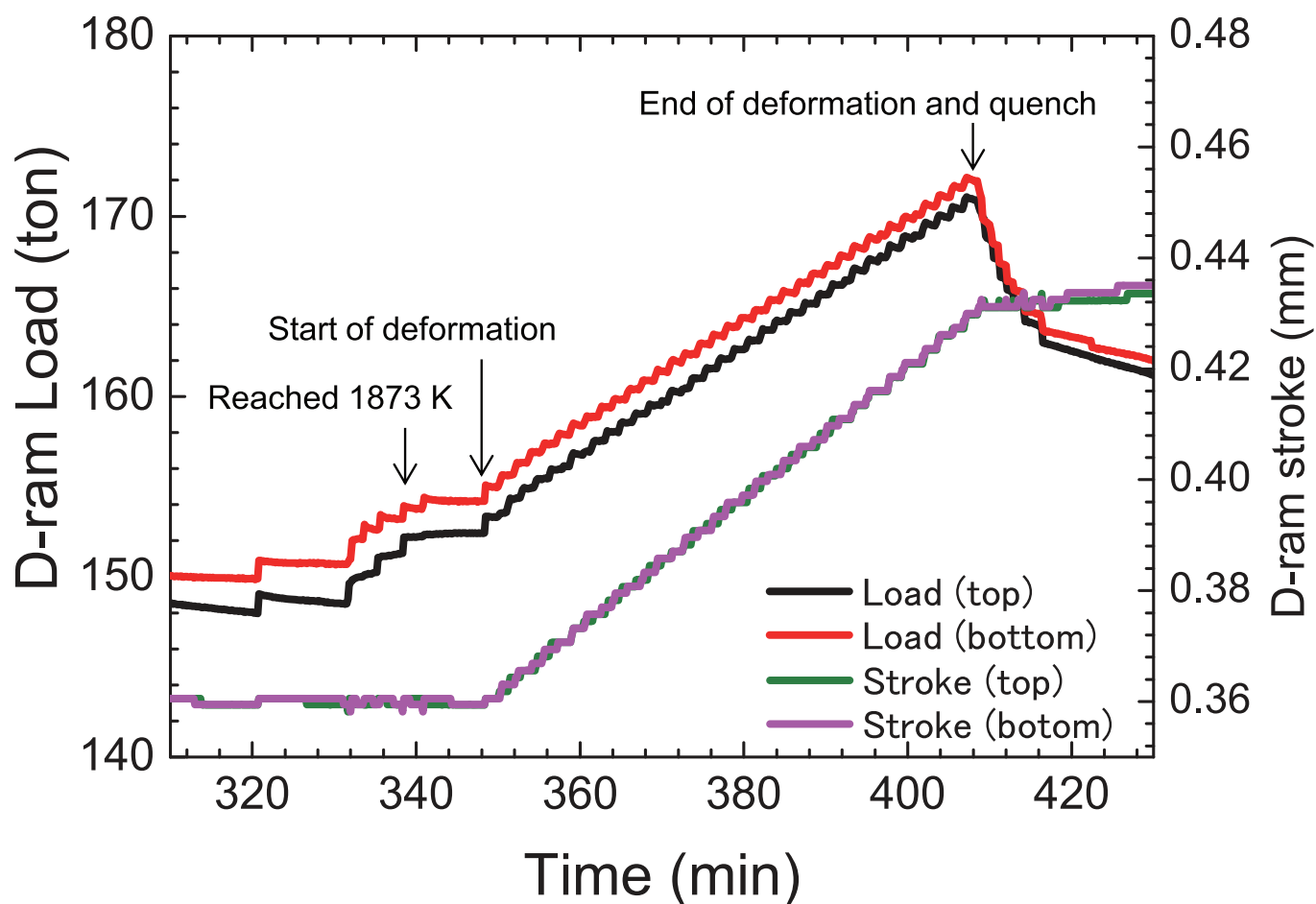
33. Yoshino, T., Yamazaki, D., Ito, E. & Katsura, T. No interconnection of ferropericlase in post-spinel phase inferred from conductivity measurement. *Geophys. Res. Lett.* **35**, L22303 (2008).
34. Rubie, D. C., Karato, S., Yan, H. & O'Neill, H. St. C. Low differential stress and controlled chemical environment in multianvil high-pressure experiments. *Phys. Chem. Miner.* **20**, 315–322 (1993).
35. Nishihara, Y. Recent technical developments of high-pressure deformation experiments. *Rev. High Press. Sci. Technol.* **18**, 223–229 (2008).
36. Morishima, H. *et al.* The phase boundary between α - and β - Mg_2SiO_4 determined by *in situ* X-ray observation. *Science* **265**, 1202–1203 (1994).
37. Inoue, T. *et al.* The phase boundary between wadsleyite and ringwoodite in Mg_2SiO_4 determined by *in situ* X-ray diffraction. *Phys. Chem. Miner.* **33**, 106–114 (2006).
38. Fei, Y. *et al.* A critical evaluation of pressure scale at high temperatures by *in situ* X-ray diffraction measurements. *Phys. Earth Planet. Inter.* **143–144**, 515–526 (2004).
39. Bolfan-Casanova, N., Keppler, H. & Rubie, D. C. Water partitioning between nominally anhydrous minerals in the MgO – SiO_2 – H_2O system up to 24 GPa: implications for the distribution of water in the Earth's mantle. *Earth Planet. Sci. Lett.* **182**, 209–221 (2000).
40. Nishihara, Y. *et al.* Stress measurement under high-pressure using Kawai-type multi-anvil apparatus combined with synchrotron radiation. *J. Synchrotron Radiat.* **16**, 757–761 (2009).
41. Ohuchi, T. *et al.* *In situ* observation of crystallographic preferred orientation of deforming olivine at high pressure and high temperature. *Phys. Earth Planet. Inter.* **243**, 1–21 (2015).
42. Niwa, K. *et al.* *In situ* observation of shear stress-induced perovskite to post-perovskite phase transition in CaIrO_3 and the development of its deformation texture in a diamond-anvil cell up to 30 GPa. *Phys. Earth Planet. Inter.* **194–195**, 10–17 (2012).



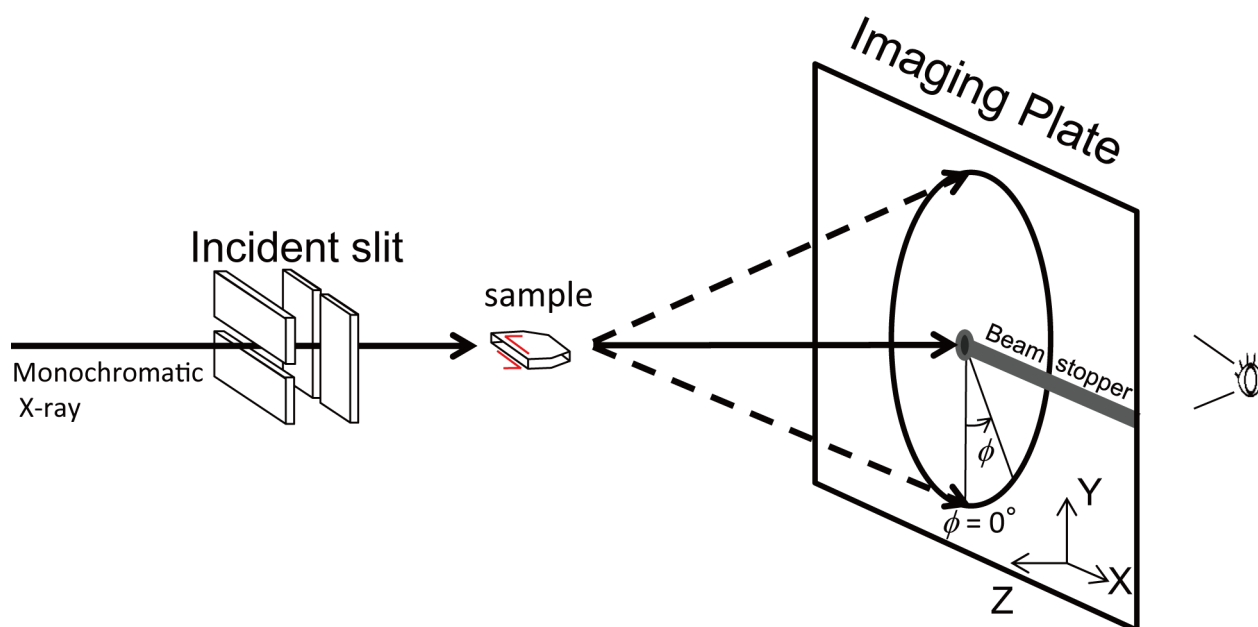
Extended Data Figure 1 | Schematic cross-sections of cell assemblies. **a**, Shear deformation experiments. **b**, Temperature calibration experiment.



Extended Data Figure 2 | The temperature and power relationship at 25 GPa using the cell assembly for the temperature calibration experiments. The results for the two runs are shown by black and red lines.

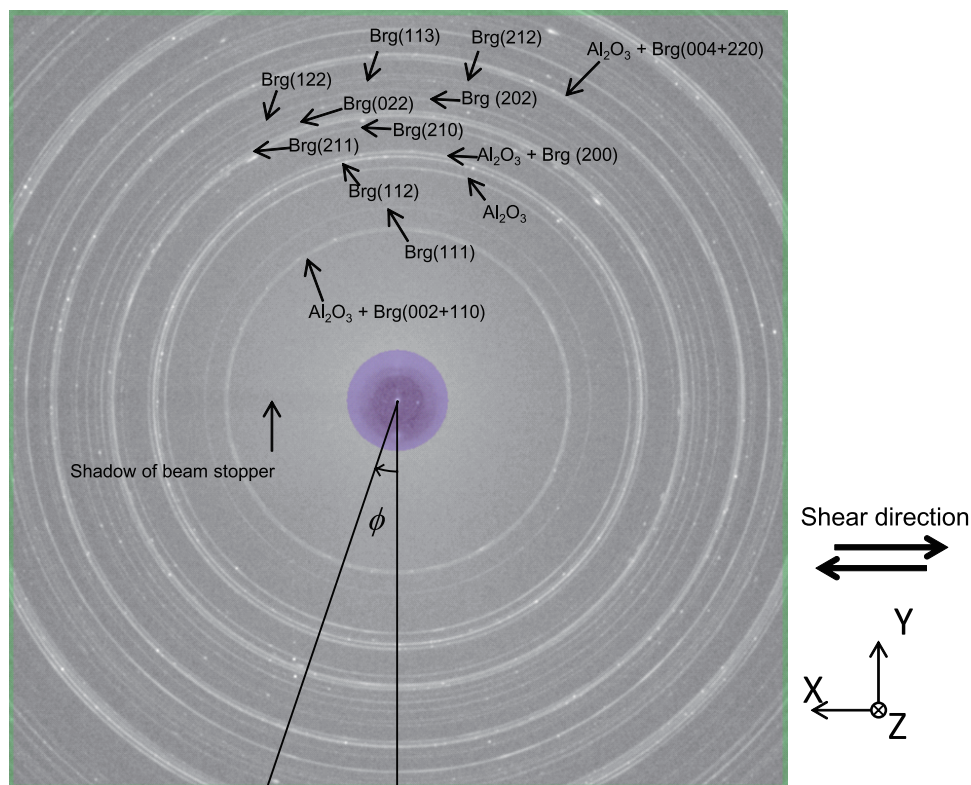


Extended Data Figure 3 | Loads and strokes of the differential rams during the deformation experiment (run K122) are shown as functions of elapsed time. The lines for the stroke of the top and bottom rams almost overlap. The sample was deformed by advancement of both the top and bottom differential rams linearly with time. During deformation, the loads of differential rams also increased linearly.

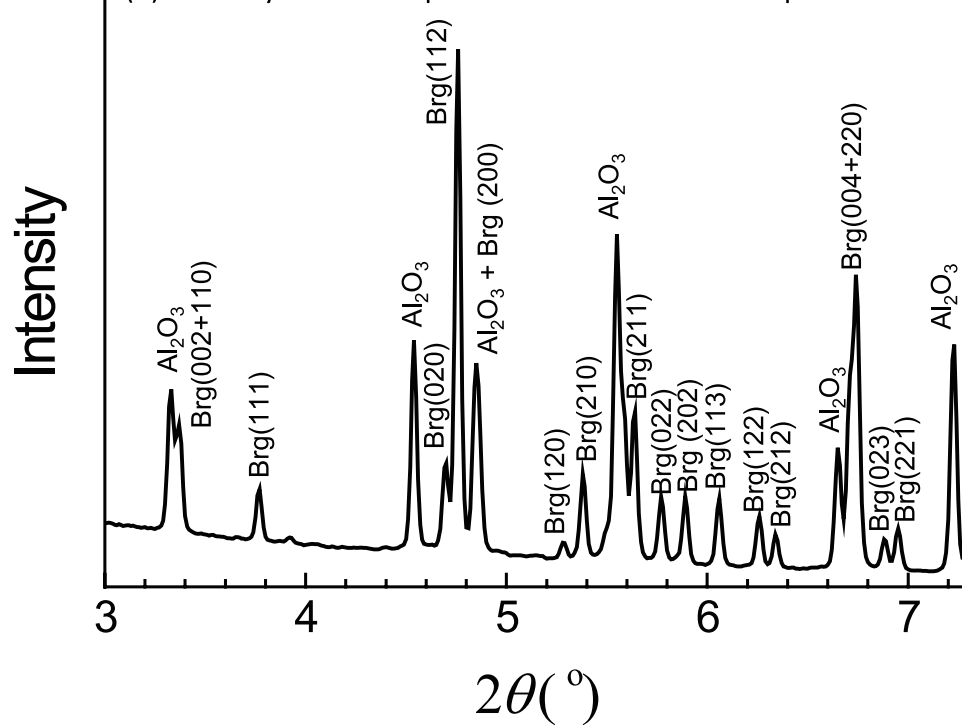


Extended Data Figure 4 | Schematic of the two-dimensional X-ray diffraction measurement set-up. For these measurements the direct beam stopper arm was located at the right side when the assembly is viewed from the front.

(a) 2D X-ray diffraction pattern of deformation sample

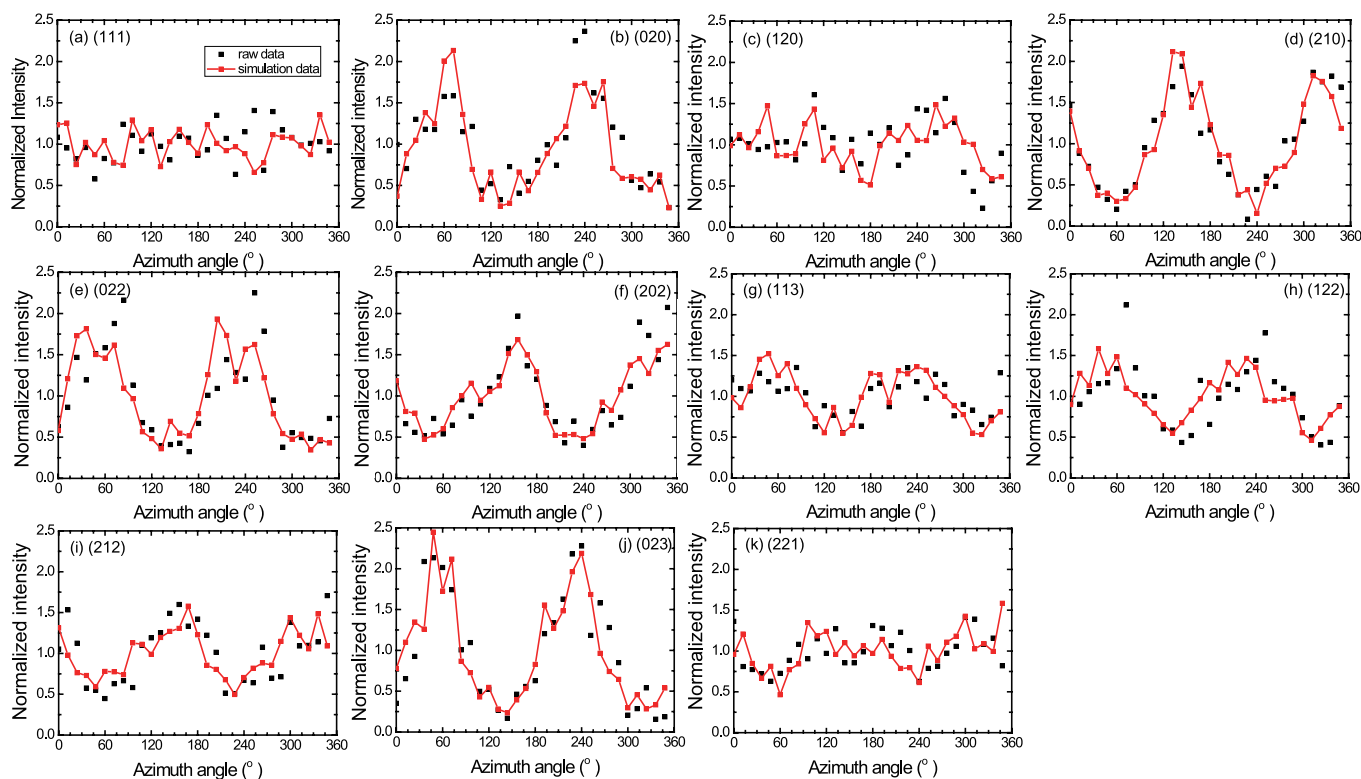


(b) 1D X-ray diffraction pattern of deformation sample

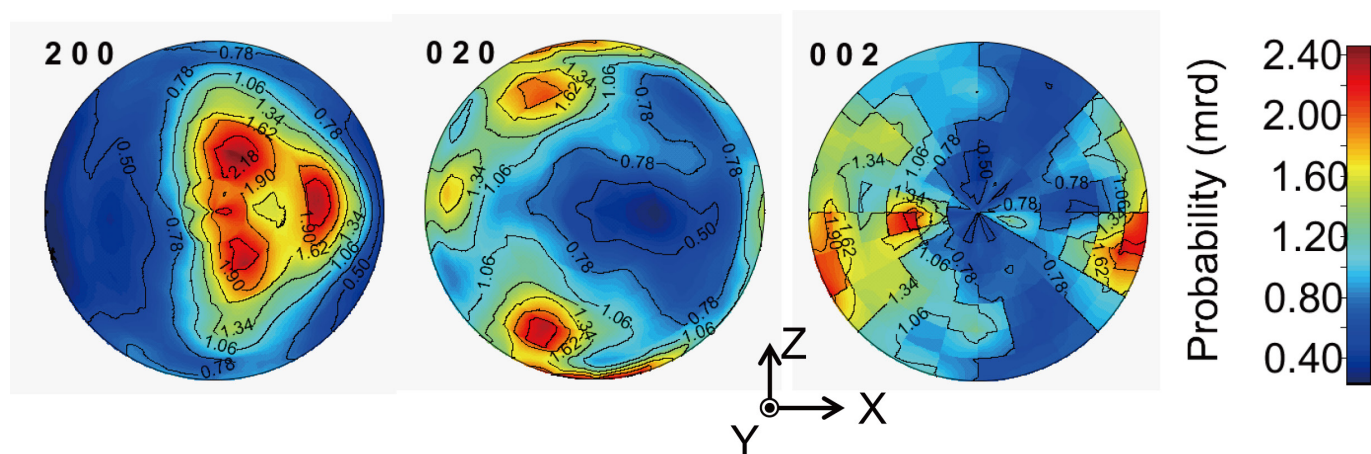


Extended Data Figure 5 | X-ray diffraction pattern of deformed bridgmanite. **a**, Two-dimensional X-ray diffraction pattern where ϕ is the azimuthal angle. **b**, One-dimensional X-ray diffraction pattern of the deformed bridgmanite (K122). In **a** the two-dimensional X-ray diffraction pattern at the imaging plate is taken from the back surface of the sample,

which is opposite to the polished plane. The position of the direct beam stopper arm is thus on the left. The diffraction peaks of the (111), (020), (120), (210), (022), (202), (113), (122), (212), (023) and (221) planes were used in the analysis of the CPO patterns. Brg, bridgmanite.



Extended Data Figure 6 | Normalized intensities of diffraction peaks of the deformed bridgmanite. Each diffraction peak intensity was normalized by its average value. a–k, Diffraction peaks for the (111) (a), (020) (b), (120) (c), (210) (d), (022) (e), (202) (f), (113) (g), (122) (h), (212) (i), (023) (j) and (221) (k) planes.



Extended Data Figure 7 | Pole figures of bridgmanite showing the variation of the CPO of the [200], [020] and [002] directions determined using MAUD. mrd, multiples of a random distribution.

Extended Data Table 1 | Summary of the dominant slip system found for bridgmanite in various studies

	Dominant slip system	Composition	Pressure	Temperature	Methods
This study	[001](100)	(Mg _{0.97} Fe _{0.03})SiO ₃	25 GPa	1873 K	Shear deformation
Merkel <i>et al.</i> (2003) ⁷	Not observed	MgSiO ₃	< 32 GPa	300 K	Uniaxial deformation by DAC
Miyagi <i>et al.</i> (2016) ²²	[100],[010] and <110> on (001) below 55 GPa, (100) plane over 55 GPa	(Mg,Fe)SiO ₃ with (Mg,Fe)O	<65 GPa	300 K	Uniaxial deformation by DAC
Cordier <i>et al.</i> (2004) ⁸	[100]	MgSiO ₃	25 GPa	300 K - 1673 K	USRE
Mainprice <i>et al.</i> (2008) ¹⁸	[010](100)	MgSiO ₃	30 GPa	0 K	FP with VPSC
Ferré <i>et al.</i> (2007) ²³	[100](010), [010](100)	MgSiO ₃	30 GPa	0 K	FP

USRE, Uniaxial stress relaxation experiments; FP, first principles; VPSC, visco-plastic self-consistent.

Extended Data Table 2 | Elasticity (C_{ij}) of the deformed bridgmanite aggregate at 25 GPa and 1,873 K

i	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
1	557.6	218.1	219.9	-0.5	0.2	2.8
2		567.7	221.4	-0.4	-0.1	3.2
3			572.2	-0.6	0.3	0.5
4				174.4	1.3	0.1
5					171.7	0.2
6						171.8

All values are in GPa. Parameters i and j indicate the reference axes: 1 corresponds to y (shear plane normal), 2 to x (shear direction) and 3 to z (perpendicular to both 1 and 2) in Fig. 2.

Extended Data Table 3 | Elasticity (C_{ij}) of a single crystal of a bridgmanite aggregate used to calculate the elasticity of deformed bridgmanite aggregate

i	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
1	539	220	210	0	0	0
2		595	232	0	0	0
3			561	0	0	0
4				187	0	0
5					178	0
6						156

The data for the single crystal of bridgmanite are from ref. 19. All values are in GPa. Parameters i and j indicate the reference axes: 1 corresponds to [100], 2 corresponds to [010] and 3 corresponds to [001] in space group $Pbnm$.

Wild monkeys flake stone tools

Tomos Proffitt^{1*}, Lydia V. Luncz^{1*}, Tiago Falótico², Eduardo B. Ottoni², Ignacio de la Torre³ & Michael Haslam¹

Our understanding of the emergence of technology shapes how we view the origins of humanity^{1,2}. Sharp-edged stone flakes, struck from larger cores, are the primary evidence for the earliest stone technology³. Here we show that wild bearded capuchin monkeys (*Sapajus libidinosus*) in Brazil deliberately break stones, unintentionally producing recurrent, conchoidally fractured, sharp-edged flakes and cores that have the characteristics and morphology of intentionally produced hominin tools. The production of archaeologically visible cores and flakes is therefore no longer unique to the human lineage, providing a comparative perspective on the emergence of lithic technology. This discovery adds an additional dimension to interpretations of the human Palaeolithic record, the possible function of early stone tools, and the cognitive requirements for the emergence of stone flaking.

Palaeoanthropologists use the distinctive characteristics of flaked stone tools both to distinguish them from naturally broken stones and to interpret the behaviour of the hominins that produced them⁴. Suggested hallmarks of the earliest stone tool technology include (i) controlled, conchoidal flaking⁵; (ii) production of sharp cutting edges⁶; (iii) repeated removal of multiple flakes from a single core; (iv) clear targeting of core edges; and (v) adoption of specific flaking patterns⁷. These characteristics underlie the identification of intentional stone flaking at all early archaeological sites^{3,5,7–12}, as they do not co-occur under natural geological conditions.

To date, comparisons between hominin intentional stone flaking and wild primate stone tool use have focused on West African chimpanzees (*Pan troglodytes verus*)^{13–16}. Nevertheless, stone breakage during chimpanzee tool use is accidental¹⁵, a result of missed hits or indirect force application during activities such as nut-cracking. The resulting stone fragments lack most of the diagnostic criteria listed above for hominin flakes^{10,17}. Even when the manufacture of sharp edges was taught to captive bonobos (*Pan paniscus*), the resulting flaked assemblage did not replicate the early hominin archaeological record¹⁸.

The capuchins of Serra da Capivara National Park (SCNP) in Brazil use stone tools in more varied activities than any other known non-human primate, including for pounding foods, digging and in sexual displays^{19–21}. Bearded capuchins and some Japanese macaques (*Macaca fuscata*) are known to pound stones directly against each other²², but the SCNP capuchins are the only wild primates that do so for the purpose of damaging those stones¹⁹. This activity, which we term stone on stone (SoS) percussion, typically involves an individual selecting rounded quartzite cobbles from a conglomerate bed (active hammers), and with one or two hands striking the hammer-stone forcefully and repeatedly on quartzite cobbles embedded within the conglomerate (passive hammers) (Fig. 1, Supplementary Video 1).

Previous observations of capuchin stone percussion indicate that this behaviour occurs in an aggressive context²³. In our observations, however, the monkeys licked or sniffed the crushed passive hammers in about half of the SoS percussion events¹⁹ (Supplementary Video 1), suggesting that they may be ingesting either powdered quartz or lichens. While the stones do not contain any biologically active components¹⁹,

silicon is known to be an essential trace nutrient²⁴. SCNP capuchins have also been seen to use a stone hammer to dislodge another stone from the conglomerate, with the second stone then used as a hammer for SoS percussion²⁰.

As well as deliberately crushing the surface of both the active and passive hammers, the capuchins regularly unintentionally fracture the stones during use (Supplementary Video 1). In addition, we observed a capuchin place a newly fractured stone flake on top of another stone, and then strike it with a hammer in a manner resembling chimpanzee nut-cracking or human bipolar reduction (Supplementary Video 1). Nevertheless, while the monkeys were seen to re-use broken hammer-stone parts as fresh hammers, they were not observed using the sharp edges of fractured tools to cut or scrape other objects.

We collected fragmented stones immediately after capuchins were observed using them at the Oitenta site in SCNP (8° 52.394' S, 42° 37.971' W) (Fig. 1), as well as from surface surveys and archaeological excavation in the same area (Extended Data Fig. 1). The assemblage consists of 111 capuchin-modified stone artefacts, including complete



Figure 1 | Wild bearded capuchin SoS percussion, Serra da Capivara National Park, Brazil. **a**, The conglomerate outcrop where SoS percussive behaviour of **b** and **c** was observed. **b**, **c**, SoS percussive actions including close observation by a juvenile capuchin (**b**), and stone breakage (**c**). Note that the active hammer in use is part of Refit Set 6 (Supplementary Information and Supplementary Video 1).

¹Primate Archaeology Research Group, School of Archaeology, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, UK. ²Institute of Psychology, University of São Paulo, São Paulo, SP 05508-030, Brazil. ³Institute of Archaeology, University College London, 31–34 Gordon Square, London WC1H 0PY, UK.

*These authors contributed equally to this work.

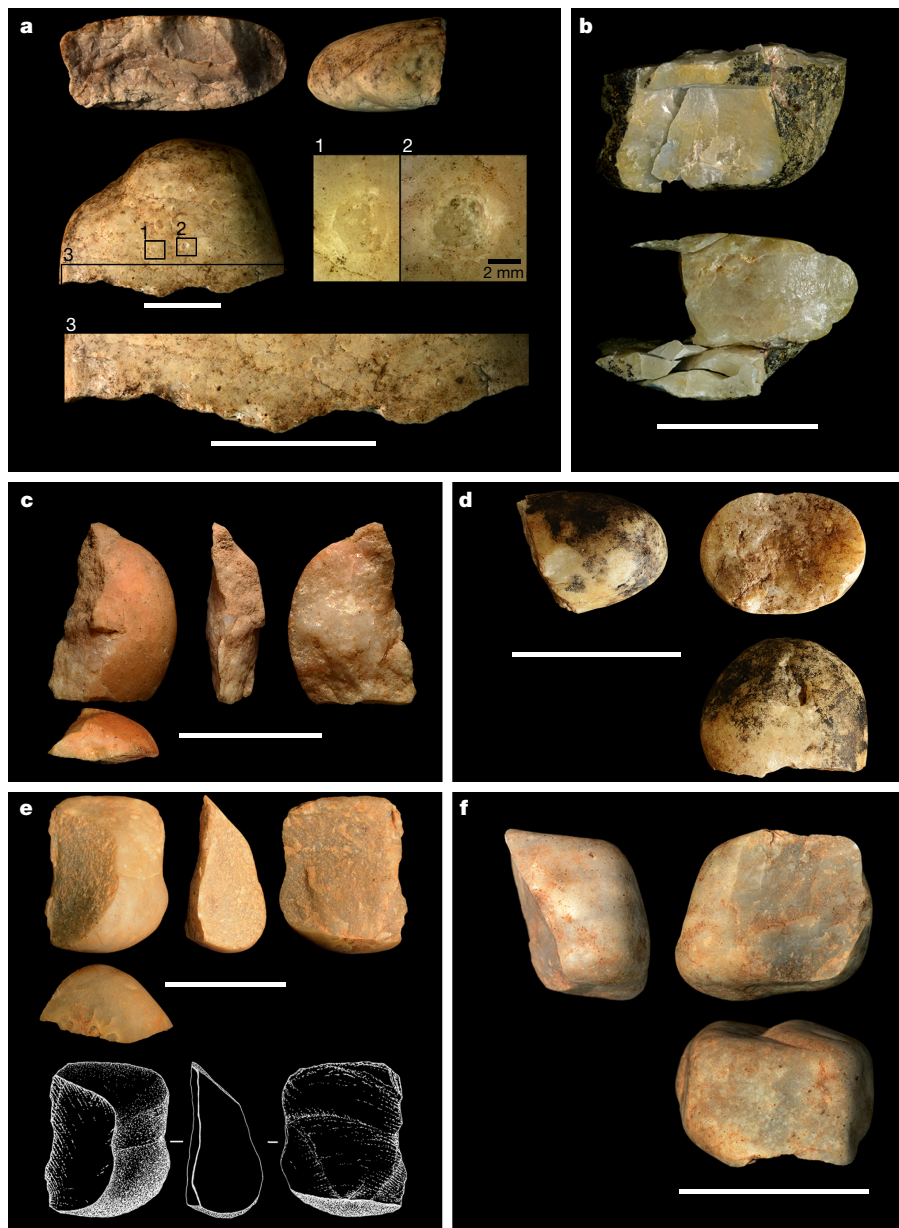


Figure 2 | Examples of flaked stones from capuchin SoS percussion. **a**, Detail of a large, unidirectionally flaked active hammer-stone, with clear impact marks located towards the centre of the striking platform. **b**, Refitted active hammer illustrating recurrent unidirectional removal of at least seven flakes (Refit Set 6; Extended Data Fig. 6b and Supplementary

Video 2). **c**, **e**, Examples of conchoidal flakes. Artefact illustrations in **e** reproduced with permission from A. Theodoropoulou. **d**, **f**, Examples of flaked hammer-stones. **a**–**f**, Scale bars are 5 cm, except for the scale bar in the inset (**a**), which is 2 mm.

and broken hammer-stones, complete and fragmented flakes, and passive hammers. We also found flaked hammer-stones, which using a traditional classification would be considered flaked artefacts²⁵ (Extended Data Table 1). All stones were originally obtained by the capuchins from conglomerates in the vicinity of their use.

Complete hammer-stones have a mean weight of 600.3 g (Extended Data Table 2a). They possess varying degrees of percussive damage across their surfaces, including small impact points surrounded by circular or crescent scars (Supplementary Information and Extended Data Fig. 2). Broken hammer-stones and flaked hammer-stones comprise over a quarter of the total assemblage. Broken hammer-stones are on average smaller than complete hammer-stones (mean = 203.8 g; Extended Data Table 2a), and some would be termed split cobbles in a hominin assemblage. Flaked hammer-stones exhibit one or more conchoidal or wedge flake scars, occurring either as 1–2 fortuitous scars from a natural striking platform, or as recurring unidirectional,

overlapping flakes resulting from repeated strikes on a fracture plane (Fig. 2, Supplementary Information and Extended Data Fig. 3). Refitted hammer-stones demonstrate this reduction sequence (Supplementary Information and Extended Data Figs 4, 5). Continuous rotation and manipulation of the hammer-stones during use also produces small (<1 cm), non-invasive, step-terminating flake scars along the edge of the striking platform, perpendicular to the flaking surface. These artefacts are indistinguishable from some archaeological examples of intentionally flaked early hominin stone cores. Using a traditional classification, the flaked hammer-stones fall within the morphology of unifacial choppers¹.

Complete flakes produced during SoS percussion have sharp edges, bulbs of percussion and scars from up to three previous flake removals (Fig. 2, Supplementary Information and Extended Data Fig. 6). A high proportion of wedge-initiated flakes occur in the early stages of reduction, evidenced by an increased frequency of cortical

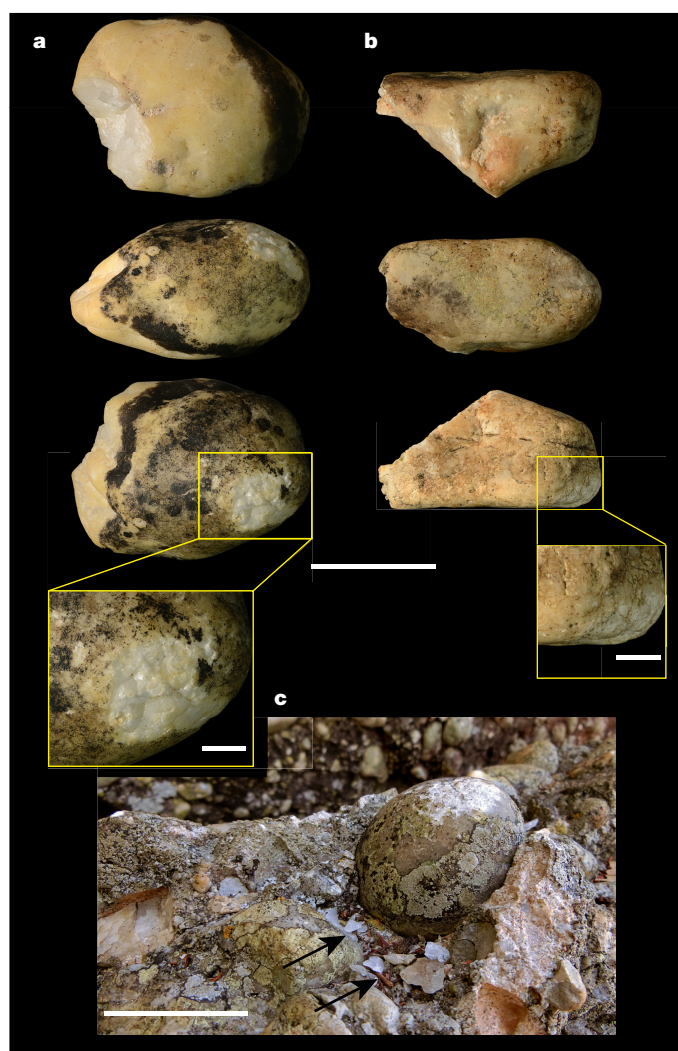


Figure 3 | Examples of passive hammers from capuchin SoS percussion. **a, b**, Passive hammers with detail of percussive damage (inset). **c**, Passive hammer *in situ* at Serra da Capivara National Park, after its observed use for SoS percussive behaviour. Note the small flake fragments at the base of the passive element, resulting from active hammer flaking. **a–c**, Main scale bars are 5 cm, the scale bars in the insets (**a, b**) are 1 cm.

flakes. Conchoidal flakes, on the other hand, come from both early and later stages of reduction, with both cortical and non-cortical pieces represented. Extensive refits record the production of unidirectional recurrent, conchoidal flakes following an initial forceful fracture (Extended Data Figs 5, 6, Supplementary Information and Supplementary Video 2).

Passive hammers, whether found detached from or embedded in the conglomerate, typically have a localized area of percussive damage located on a prominent surface (Fig. 3). The damage includes impact points, battering marks and crushed quartz crystals and, in some cases, detached flakes or chips. The passive hammers in this study (mean = 303.7 g, Extended Data Table 2a) also retain evidence of their subsequent re-use as active hammers, with impact points located on previously embedded flat planes opposite the passive hammer damage. This use clearly occurred after the stone was dislodged from the conglomerate. Capuchin SoS tools are therefore multifunctional, with the monkeys able to repurpose stones from a passive to an active percussive role (Supplementary Information).

The distinctive assemblages found at SoS percussion sites will guide future archaeological investigations into the development of capuchin technology at SCNP²⁶, and the broader Middle Pleistocene dispersal of *Sapajus* into northeast Brazil²⁷. They should also assist in distinguishing

human tools from capuchin artefacts where the ranges of these primates overlap¹². Of interest beyond *Sapajus* behavioural evolution, SCNP capuchins produce stone debris through a similar technique (passive hammer) to that inferred from some of the earliest hominin archaeological assemblages^{3,11}. The passive hammer knapping technique involves striking a hammer-stone onto a passive anvil, with the desired flakes detached from the hand-held stone¹¹ (Supplementary Video 1). Both active and passive hominin hammers often have repeated impact marks away from the tool's edge, interpreted as evidence of poorly controlled strikes or multi-purpose tool use³. SCNP capuchin behaviour demonstrates that these marks and recurrent conchoidally fractured, sharp-edged flakes, can be produced entirely unintentionally.

The SCNP data provide an example of repeated conchoidal flaking that is not reliant on advanced, human-like hand morphologies and coordination²⁸. Similarly, SoS behaviour presents an alternative to evolutionary explanations that link the origins of recurrent flake production to a change in hominin cognitive skills^{28,29}. In the absence of supporting evidence such as cut-marked bones, we suggest that sharp-edged flake production can no longer be implicitly or solely associated with intentional production of cutting flakes. Capuchin SoS percussion and simple Pliocene–Pleistocene stone knapping activities are equifinal behaviours in the production of flaked lithic assemblages. These findings open up the possibility that unintentional flaked assemblages may be identified in the palaeontological record of extinct apes and monkeys. In light of this possibility, criteria commonly used to distinguish intentional hominin lithic assemblages need to be refined.

No living primate is a direct substitute for extinct hominins, which varied in unknown ways from the behaviour, cognition and morphology seen in extant animals and humans¹⁵. However, capuchin SoS percussion is an example of intentional stone breakage by a non-human primate that produces concentrated lithic accumulations. Capuchin SoS percussion flakes and flaked hammer-stones fall within the range of mean dimensions for simple flakes and cores from the Early Stone Age³ (Supplementary Information and Extended Data Table 2b). If encountered in a hominin archaeological context, this material would be identified as artefactual, potentially interpreted as the result of intentional stone fracture and controlled flake production, and probably attributed to functional needs requiring the use of sharp edges.

The capuchin data add support to an ongoing paradigm shift in our understanding of stone tool production and the uniqueness of hominin technology. Within the last decade, studies have shown that the use³⁰ and intentional production³ of sharp-edged flakes is not necessarily tied to the genus *Homo*. Capuchin SoS percussion goes a step further, demonstrating that the production of archaeologically identifiable flakes and cores, as currently defined, is no longer unique to the human lineage.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 June; accepted 21 September 2016.

Published online 19 October 2016.

1. Leakey, M. D. *Olduvai Gorge, Vol. 3. Excavations in Beds I and II, 1960–1963*. (Cambridge University Press, 1971).
2. Potts, R. *Early hominid activities at Olduvai*. (AldineTransaction, 1988).
3. Harmand, S. *et al.* 3.3-million-year-old stone tools from Lomekwi 3, West Turkana, Kenya. *Nature* **521**, 310–315 (2015).
4. Panger, M. A., Brooks, A. S., Richmond, B. G. & Wood, B. Older than the Oldowan? Rethinking the emergence of hominin tool use. *Evol. Anthr.* **11**, 235–245 (2002).
5. Isaac, G. L. I. Early hominids in action: a commentary on the contribution of archaeology to understand the fossil record in East Africa for 1975. *Yrbk. Phys. Anthr.* **19–35** (1976).
6. Toth, N. The Oldowan reassessed: a close look at early stone artifacts. *J. Archaeol. Sci.* **12**, 101–120 (1985).
7. Delagnes, A. & Roche, H. Late Pliocene hominid knapping skills: the case of Lokalalei 2C, West Turkana, Kenya. *J. Hum. Evol.* **48**, 435–472 (2005).

8. Semaw, S. The world's oldest stone artefacts from Gona, Ethiopia: their implications for understanding stone technology and patterns of human evolution between 2.6–1.5 million years ago. *J. Archaeol. Sci.* **27**, 1197–1214 (2000).
9. Stout, D., Semaw, S., Rogers, M. J. & Cauche, D. Technological variation in the earliest Oldowan from Gona, Afar, Ethiopia. *J. Hum. Evol.* **58**, 474–491 (2010).
10. de la Torre, I. Ormo revisited: evaluating the technological skills of Pliocene hominids. *Curr. Anthropol.* **45**, 439–465 (2004).
11. Lewis, J. E. & Harmand, S. An earlier origin for stone tool making: implications for cognitive evolution and the transition to *Homo*. *Phil. Trans. R. Soc. B* **371**, 20150233 (2016).
12. Boëda, E. *et al.* A new late Pleistocene archaeological sequence in South America: the Vale da Pedra Furada (Piauí, Brazil). *Antiquity* **88**, 927–941 (2014).
13. Matsuzawa, T., Humle, T. & Sugiyama, Y. *The Chimpanzees of Bossou and Nimba*. (Springer Science & Business Media, 2011).
14. Mercader, J., Panger, M. & Boesch, C. Excavation of a chimpanzee stone tool site in the African rainforest. *Science* **296**, 1452–1455 (2002).
15. McGrew, W. C. *Chimpanzee Material Culture: Implications for Human Evolution*. (Cambridge Univ. Press, 1992).
16. Mercader, J. *et al.* 4,300-year-old chimpanzee sites and the origins of percussive stone technology. *Proc. Natl Acad. Sci. USA* **104**, 3043–3048 (2007).
17. Pelegrin, J. in *Stone Knapping: The Necessary Conditions for a Uniquely Hominid Behaviour*. (eds Roux, V. & Bril, B.) 23–33 (McDonald Institute monograph series, 2005).
18. Toth, N., Schick, K. & Semaw, S. in *The Oldowan: Case Studies into the Earliest Stone Age* (eds Toth, N. & Schick, K.) 155–222 (Stone Age Institute Press, 2006).
19. Falótico, T. & Ottoni, E. B. The manifold use of pounding stone tools by wild capuchin monkeys of Serra da Capivara National Park, Brazil. *Behaviour* **153**, 421–442 (2016).
20. Mannu, M. & Ottoni, E. B. The enhanced tool-kit of two groups of wild bearded capuchin monkeys in the Caatinga: tool making, associative use, and secondary tools. *Am. J. Primatol.* **71**, 242–251 (2009).
21. Falótico, T. & Ottoni, E. B. Stone throwing as a sexual display in wild female bearded capuchin monkeys, *Sapajus libidinosus*. *PLoS One* **8**, e79535 (2013).
22. Leca, J.-B., Gunst, N. & Huffman, M. Complexity in object manipulation by Japanese macaques (*Macaca fuscata*): a cross-sectional analysis of manual coordination in stone handling patterns. *J. Comp. Psychol.* **125**, 61–71 (2011).
23. Moura, A. C. Stone banging by wild capuchin monkeys: an unusual auditory display. *Folia Primatol. (Basel)* **78**, 36–45 (2007).
24. Carlisle, E. M. Silicon as an essential trace element in animal nutrition. *Silicon Biotech.* **703**, 123–139 (2008).
25. Isaac, G. L. *Koobi Fora Research Project Vol. 5: Plio-Pleistocene Archaeology*. (Clarendon, 1997).
26. Haslam, M. *et al.* Pre-Columbian monkey tools. *Curr. Biol.* **26**, R521–R522 (2016).
27. Haslam, M. Towards a prehistory of primates. *Antiquity* **86**, 299–315 (2012).
28. Kivell, T. L. Evidence in hand: recent discoveries and the early evolution of human manual manipulation. *Phil. Trans. R. Soc. B* **370**, 20150105 (2015).
29. de la Torre, I. in *Stone Tools and the Evolution of Human Cognition*. (eds Nowell, A. & Davidson, I.) 45–65 (Univ. Press of Colorado, 2010).
30. McPherron, S. P. *et al.* Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature* **466**, 857–860 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements The study was funded by a European Research Council Starting Investigator Grant (#283959) to M.H. and São Paulo Research Foundation (FAPESP) awards to T.F. (#2013/05219-0) and E.B.O. (#2014/04818-0). Support for fieldwork and analysis was provided by N. Guidon and G. Daltrini Felice of FUMDHAM and University College London (ERC Starting Grant #283366). We thank R. Fonseca de Oliveira for excavation coordination, M. Gumert, R. Mora and A. Arroyo for comments, and A. Theodoropoulou for artefact illustrations. Fieldwork at SCNP was approved by Brazilian environmental protection agencies (IBAMA/ICMBio 37615-2).

Author Contributions M.H. and T.F. observed and recorded the capuchin behaviour, collected lithic material and directed excavations at Serra da Capivara National Park. T.P. conducted the technological analysis. T.P., L.V.L., I.D.L.T. and M.H. discussed the implications of the results. T.P. wrote the paper and supplementary online content with contributions from L.V.L., T.F., E.B.O., I.D.L.T. and M.H. T.P. generated all figures, 3D models and video content, using data recorded by M.H. and T.P.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.P. (tomos.proffitt@arch.ox.ac.uk) or M.H. (michael.haslam@arch.ox.ac.uk).

Reviewer Information *Nature* thanks S. Carvalho and H. Roche for their contribution to the peer review of this work.

METHODS

The SoS percussion assemblage included 111 artefacts collected from surface and archaeological capuchin activity locations in Serra da Capivara National Park (SCNP), Piauí, Brazil. The surface collection (Lasca OIT surface; $n = 60$, 54.1%) was produced by capuchins observed performing SoS percussion in September 2014, at a site later designated Lasca Oitente 2 (Lasca OIT 2). The capuchins belong to the Jurubeba group, which was first studied in March 2004 (ref. 20). SoS activity primarily took place on a low (approximately 1 m high), narrow conglomerate ridge associated with a much larger conglomeratic outcrop (Fig. 1; Supplementary Video 1). During this time a portion of the used assemblage dropped to the ground immediately below the activity area, and was collected once the activity ceased. Additional material was collected during surface surveys within the immediate vicinity of Lasca OIT 2, at locations where isolated conglomerate blocks were used by the same capuchin group for SoS percussion. This material was also analysed as Lasca OIT surface.

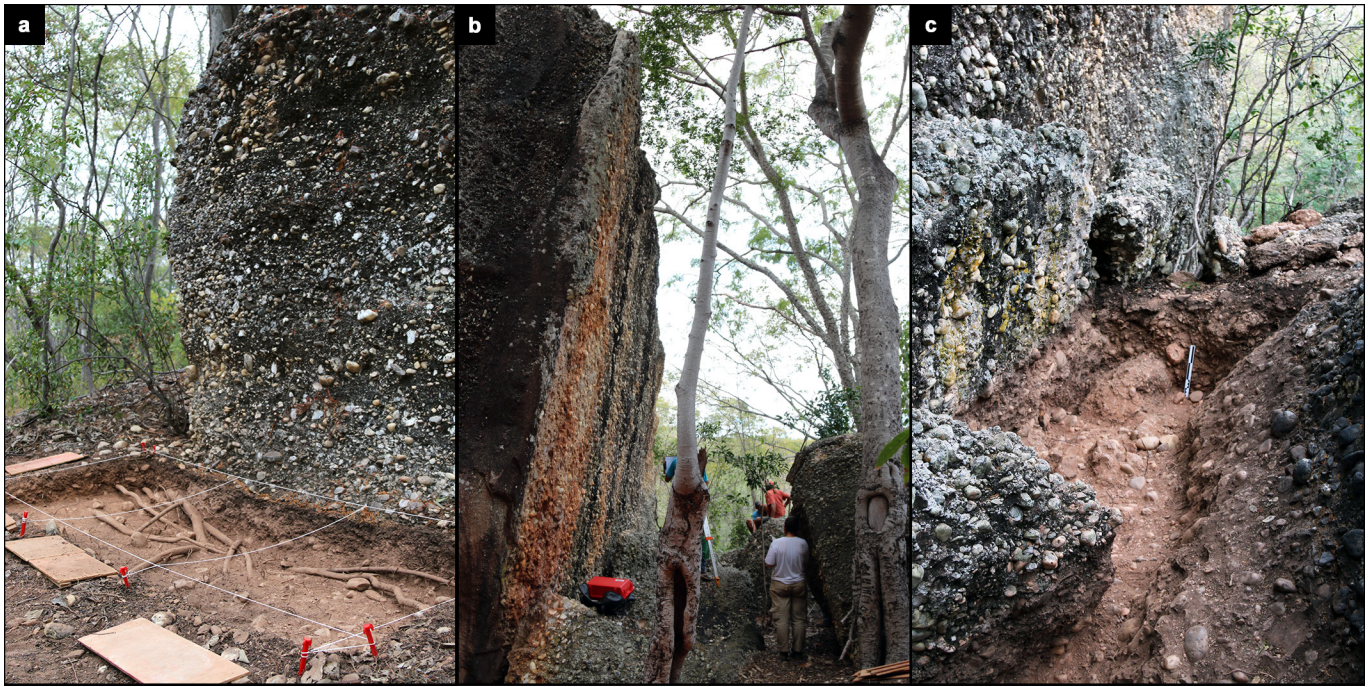
The archaeological material comes from two excavations conducted in June 2015 (Extended Data Fig. 1), within the Jurubeba group range: Lasca OIT 1 (8° 52.460' S, 42° 37.977' W) and Lasca OIT 2 (8° 52.394' S, 42° 37.971' W). We excavated both sites by hand in 5-cm levels, and sieved all sediment through a 5 mm mesh. Sediments at both sites consisted of light-brown, silty sand, with gravel to cobble-sized inclusions, resulting from the *in situ* weathering of local conglomerates. We distinguished capuchin tools from natural stones on the basis of percussion marks and flaking features as described in the main text and below. The Lasca OIT 2 excavation (Extended Data Fig. 1b) can be considered an extension of the surface material collected in 2014 from the same site. An area of 3 m² excavated to a maximum depth of 0.5 m yielded 28 SoS percussion artefacts (25.2%) at Lasca

OIT 2. We excavated Lasca OIT 1 (Extended Data Fig. 1a), located 120 m southwest of Lasca OIT 2, beneath the sheer face of an approximately 7 m high conglomerate outcrop that showed percussion marks indicative of previous SoS activity. A total excavated area of 3 m² to a maximum depth of 0.4 m yielded 23 artefacts (20.7%) at this site. We did not find human material, such as hearths, ceramic pieces, metal objects, or ground stone at either site. Such items are ubiquitous in anthropogenic sites elsewhere in SCNP³¹. This absence, along with direct observation of capuchins creating the flaked surface assemblage, and the identical nature of the damage and size of the recovered stones to those observed in use by capuchins, rules out human production of the archaeological material.

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

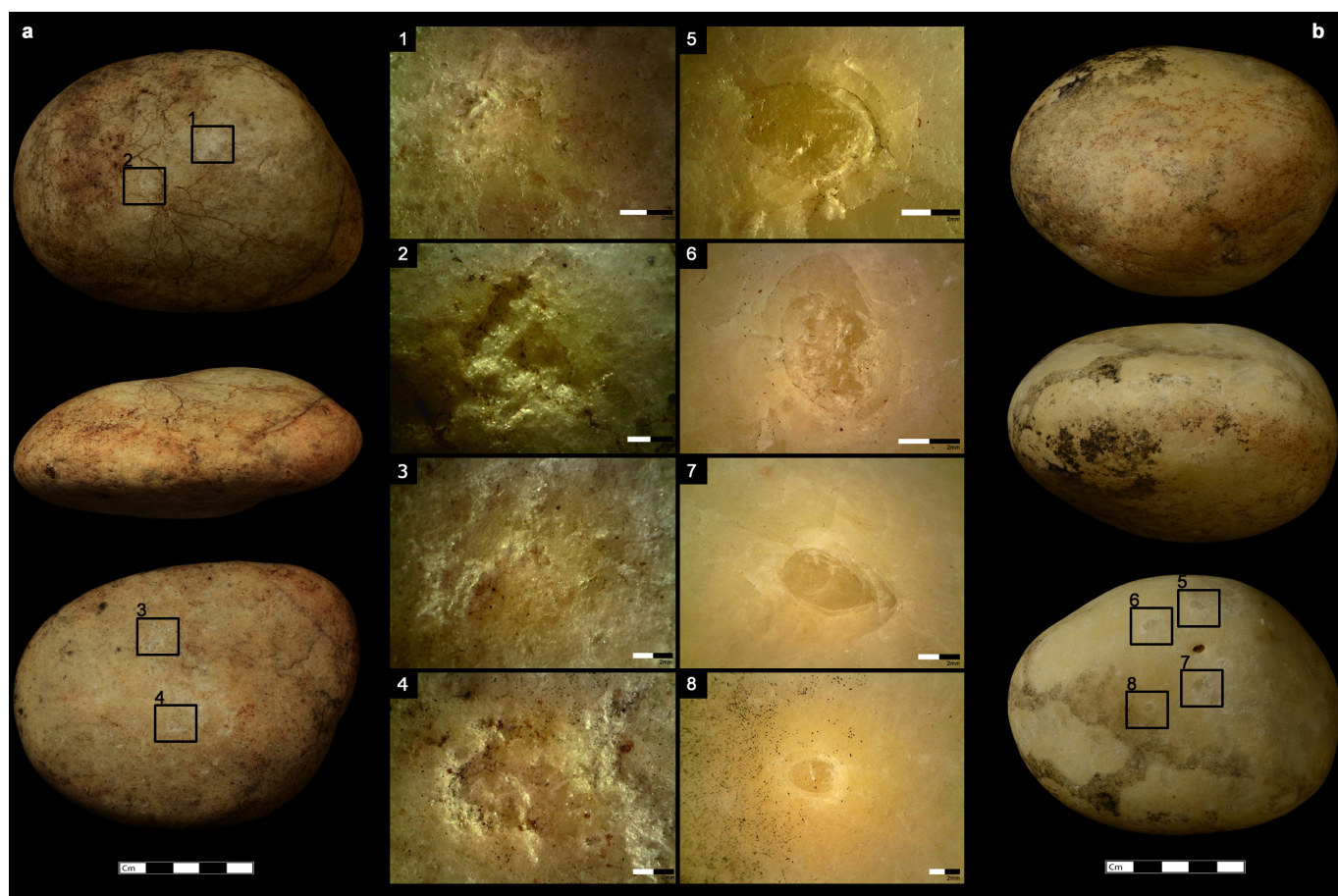
We identified the raw material of each artefact and performed technological classification and analysis following commonly used technological attributes^{7,9,32,33}. For full details and definitions of the technological categories used in this analysis, see the Supplementary Information. All data are available upon request.

31. Pessiss, A.-M., Martin, G. & Guidon, N. *Os Biomad e as Sociedades Huanas na Pre-Historia da Região do Parque nacional Serra da Capivara, Brasil*. Volume II A–B. (Fundação Museu Do Homem Americano—Fumdam, Ipsis Gráfica E Editora, 2014).
32. Inizan, M.-L., Reduron-Ballinger, M. & Roche, H. *Technology and Terminology of Knapped Stone: Followed by a Multilingual Vocabulary Arabic, English, French, German, Greek, Italian, Portuguese, Spanish*. **5**, (Cercle de Recherches et d'Etudes Préhistoriques, 1999).
33. De la Torre, I. & Mora, R. *Technological Strategies in the Lower Pleistocene at Olduvai Beds I & II*. (Univ. Liège press, ERAUL, 2005).

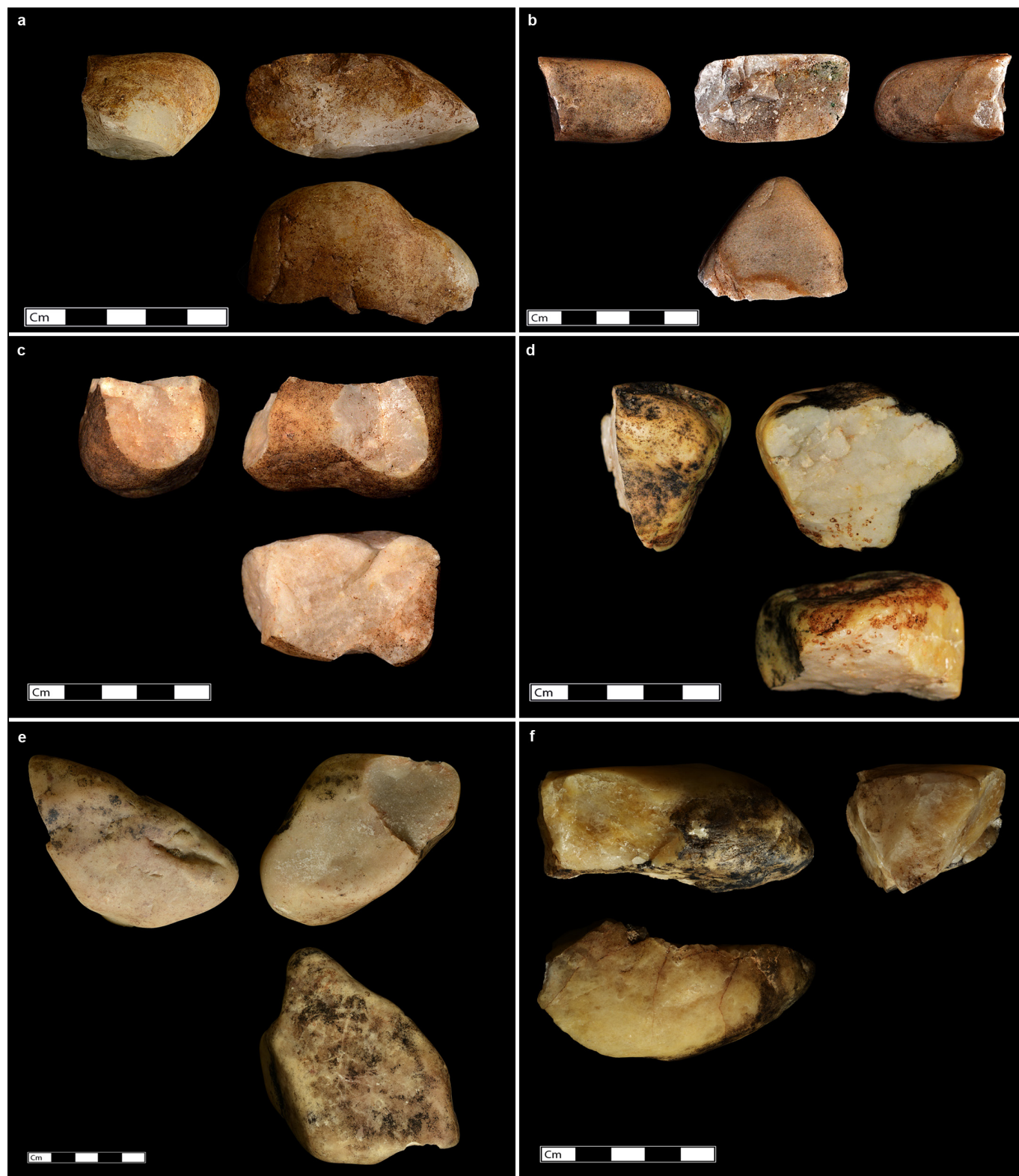


Extended Data Figure 1 | Archaeological excavation of wild capuchin SoS percussion sites, Serra da Capivara National Park. a, Lasca OIT1 excavation, each square is 1×1 m. **b,** The approach to Lasca OIT2, which is located to the right of the conglomerate cliff face. **c,** Lasca

OIT2 excavation, note the low conglomerate ridge to the left, on which capuchins were observed whilst performing SoS activities. Scale bar, 30 cm (see also Fig. 1).



Extended Data Figure 2 | Examples of active hammers. a, Crushing impacts on multiple surfaces of an active hammer. **b,** Examples of impact points and associated circular hertzian fractures on the surface of an active hammer. Scale bars are 5 cm, except for inset scale bars, which are 2 mm.

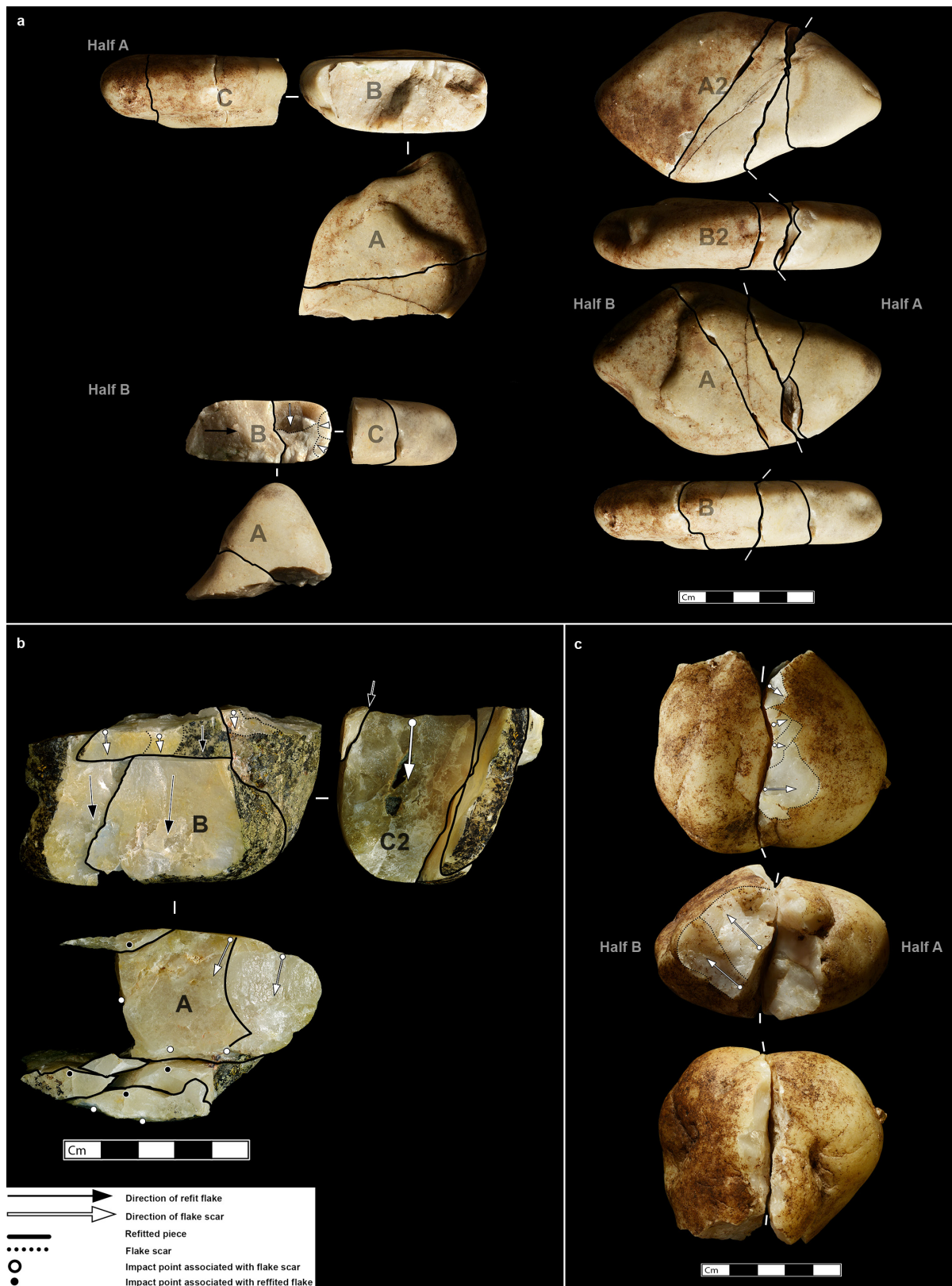


Extended Data Figure 3 | Examples of SoS flaked hammer-stones. a, c, Flake detachment following a transverse active hammer fracture. b, Unintentional radial reduction of flaked hammer-stone. d–f, Examples of complete active hammers with scars of accidental flakes. Scale bars are 5 cm.



Extended Data Figure 4 | Refits of flaked hammer-stones showing the repeated detachment of unidirectional flakes. a, Refit Set 1 (artefact numbers JC13 and JF7). **b,** Refit Set 2 (artefact numbers 225102a and 225102b). **c,** Refit Set 3 (artefact numbers 224881a and 224881b). **d,** Refit

Set 4 (artefact numbers JF3 and JC5). A, A2, B and C are designated planes on each refit, corresponding to descriptions found in Supplementary Information. Scale bars are 5 cm.



Extended Data Figure 5 | Refits of flaked hammer-stones showing the repeated detachment of unidirectional flakes and continued use of broken active hammers. a, Refit Set 5 (artefact numbers JC11, JC12, JF23 and JF1). b, Refit Set 6 (artefact numbers JC6, JF2, JF14, JF4 and JF8)

(See also Supplementary Video 2). **c, Refit Set 7 (artefact numbers JC4 and JC10). A, A2, B, B2, C and C2 are designated planes on each refit, corresponding to descriptions found in Supplementary Information. Scale bars are 5 cm.**



Extended Data Figure 6 | Examples of complete flakes. a–f, Examples of complete flakes detached during capuchin SoS percussion. Scale bars are in cm. Scale bars are 5 cm.

Extended Data Table 1 | Absolute and relative frequencies and total weights (g) of technological categories identified in each Capuchin SoS assemblage, Serra da Capivara National Park

Technological Category	Assemblage															
	Lasca OIT Surface		Lasca OIT 1 Excavation		Lasca OIT 2 Excavation		Total		Lasca OIT Surface		Lasca OIT 1 Excavation		Lasca OIT 2 Excavation		Total	
	Frequency								Total Weight (g)							
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Complete Hammerstone	0	0.0	8	34.8	8	28.6	16	14.4	0	0.0	4602	58.1	5002	61.2	9605	51.7
Broken Hammerstone	4	6.7	2	8.7	6	21.4	12	10.8	697.9	27.9	570.4	7.2	1698	23.2	3166	17.0
Flaked Pieces	10	16.7	4	17.4	7	25.0	21	18.9	1297	51.8	2154	27.2	832.5	10.2	4283	23.0
Complete Flake	19	31.7	7	30.4	5	17.9	31	27.9	288.4	11.5	167.1	2.1	179.5	2.2	635	3.4
Fragmented Flake	12	20.0	1	4.3	0	0.0	13	11.7	418	17	27	0.3	0	0.0	68.8	0.4
Chunk	13	21.7	0	0.0	1	3.6	14	12.6	178.3	7.1	0	0.0	511	0.6	229.4	1.2
Small Debris	2	3.3	0	0.0	0	0.0	2	1.8	14	0.1	0	0.0	0	0.0	14	0.0
Passive Element	0	0.0	1	4.3	1	3.6	2	1.8	0	0.0	398.3	5.0	209.1	2.6	607.4	3.3
Total	60	100	23	100	28	100	111	100	2505	100	7919	100	8172	100	18595	100

Extended Data Table 2 | Dimensional data for all artefacts from Capuchin SoS assemblages and a comparison with Pliocene–Pleistocene hominin artefacts

A	Technological Category	Measure	Assemblage															
			Lasca OT Surface				Lasca OT 1 Excavation				Lasca OT 2 Excavation				Total			
			Min	Max	Mean	StDev	Min	Max	Mean	St.Dev	Min	Max	Mean	St.Dev	Min	Max	Mean	St.Dev
Hammerstone	Max Length (mm)	-	-	-	-	75.00	129.00	101.15	19.35	61.70	149.50	104.89	27.19	61.70	149.50	103.02	22.88	
	Max Width (mm)	-	-	-	-	53.90	93.70	77.16	12.02	48.80	126.40	76.90	24.56	48.80	126.40	77.03	18.68	
	Max Thickness (mm)	-	-	-	-	39.70	74.90	54.63	11.05	40.00	74.80	54.34	14.30	39.70	74.80	54.48	12.35	
	Weight (g)	-	-	-	-	261.50	924.20	575.30	231.89	155.40	1569.00	625.28	469.54	155.40	1569.00	600.29	358.67	
Broken Hammerstone	Max Length (mm)	48.16	92.58	77.33	20.24	75.10	101.70	88.40	18.81	50.90	108.10	94.18	22.17	48.16	108.10	87.60	20.72	
	Max Width (mm)	30.91	61.66	43.65	13.92	43.60	67.10	55.35	16.62	28.00	76.20	55.25	16.29	28.00	76.20	51.40	15.21	
	Max Thickness (mm)	18.87	42.87	34.46	11.03	42.10	57.60	49.85	10.96	14.00	60.90	44.18	17.01	14.00	60.90	41.89	14.49	
	Weight (g)	37.80	286.30	174.48	102.87	187.40	383.00	285.20	138.31	26.90	604.90	316.25	194.21	26.90	604.90	263.82	162.03	
Flaked Pieces	Max Length (mm)	42.97	81.52	64.81	12.22	85.50	137.50	101.23	24.45	43.00	101.40	64.76	21.38	42.00	137.50	71.73	22.56	
	Max Width (mm)	35.52	57.01	44.98	6.94	41.30	99.50	67.08	24.52	32.00	47.00	39.53	5.48	32.00	99.50	47.37	14.93	
	Max Thickness (mm)	23.38	45.29	35.79	8.48	33.90	77.80	56.33	18.11	22.50	43.60	32.64	7.87	22.50	77.80	38.65	13.39	
	Weight (g)	48.90	247.00	129.69	53.44	136.70	1101.00	538.40	412.77	62.60	265.40	118.93	79.26	48.90	1101.00	203.95	237.45	
Complete Flakes	Max Length (mm)	14.80	70.98	34.95	16.69	29.90	63.60	44.34	11.17	30.20	65.80	49.70	12.72	14.80	70.98	39.45	15.80	
	Max Width (mm)	6.00	45.70	22.48	11.72	23.70	44.70	31.54	8.02	16.50	52.40	34.66	15.39	6.00	52.40	26.49	12.42	
	Max Thickness (mm)	1.60	27.56	12.03	7.34	6.90	24.90	13.57	6.54	11.10	33.50	16.94	9.32	1.60	33.50	13.17	7.46	
	Weight (g)	.10	44.70	15.18	16.15	4.80	59.20	23.87	23.79	9.90	108.60	35.90	41.51	.10	108.60	20.48	23.65	
Fragmented Flakes	Max Length (mm)	13.04	41.75	21.32	9.49	79.30	79.30	79.30	-	-	-	-	-	13.04	79.30	25.78	18.47	
	Max Width (mm)	7.05	24.42	13.37	4.93	28.60	28.60	28.60	-	-	-	-	-	7.05	28.60	14.54	6.33	
	Max Thickness (mm)	3.71	19.92	6.96	4.64	14.60	14.60	14.60	-	-	-	-	-	3.71	19.92	7.55	4.93	
	Weight (g)	.40	26.00	3.48	7.18	27.00	27.00	27.00	-	-	-	-	-	.40	27.00	5.29	9.48	
Chunk	Max Length (mm)	12.63	57.31	26.74	13.33	-	-	-	-	70.40	70.40	70.40	-	12.63	70.40	29.86	17.32	
	Max Width (mm)	9.89	55.32	19.48	12.39	-	-	-	-	38.00	38.00	38.00	-	9.89	55.32	20.81	12.89	
	Max Thickness (mm)	7.01	33.54	13.60	7.38	-	-	-	-	25.00	25.00	25.00	-	7.01	33.54	14.41	7.72	
	Weight (g)	.70	83.00	13.72	24.10	-	-	-	-	51.10	51.10	51.10	-	.70	83.00	16.39	25.22	
Small Debris	Max Length (mm)	14.03	14.38	14.21	.25	-	-	-	-	-	-	-	-	14.03	14.38	14.21	.25	
	Max Width (mm)	6.22	8.36	7.29	1.51	-	-	-	-	-	-	-	-	6.22	8.36	7.29	1.51	
	Max Thickness (mm)	5.23	8.11	6.67	2.04	-	-	-	-	-	-	-	-	5.23	8.11	6.67	2.04	
	Weight (g)	.60	.80	.70	.14	-	-	-	-	-	-	-	-	.60	.80	.70	.14	
Passive Hammers	Max Length (mm)	-	-	-	-	90.20	90.20	90.20	-	85.3	85.3	85.3	-	85.30	90.20	87.75	3.46	
	Max Width (mm)	-	-	-	-	68.70	68.70	68.70	-	53.7	53.7	53.7	-	53.70	68.70	61.20	10.61	
	Max Thickness (mm)	-	-	-	-	50.50	50.50	50.50	-	38.1	38.1	38.1	-	38.10	50.50	44.30	8.77	
	Weight (g)	-	-	-	-	398.30	398.30	398.30	-	209.1	209.1	209.1	-	209.10	398.30	303.70	133.78	
Natural/Unmodified	Max Length (mm)	-	-	-	-	-	-	-	-	84.20	93.90	89.05	6.86	84.20	93.90	89.05	6.86	
	Max Width (mm)	-	-	-	-	-	-	-	-	78.30	82.00	80.15	2.62	78.30	82.00	80.15	2.62	
	Max Thickness (mm)	-	-	-	-	-	-	-	-	35.70	44.10	39.90	5.94	35.70	44.10	39.90	5.94	
	Weight (g)	-	-	-	-	-	-	-	-	271.50	463.40	367.45	135.69	271.50	463.40	367.45	135.69	
B	Site	Age (Ma)	Length (mm)				Width (mm)				Thickness (mm)							
			N	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max			
	Flakes																	
	LOM3	3.3	26	120.0	48.80	19	205	110.1	40.70	19	185	43.9	23.40	6	90			
	OGS7	2.6	73	39.1	14.30	13	80	37.1	14.10	13	74	12.7	5.07	3	26			
	EG10	2.6	114	37.4	15.34	14	78	34.6	13.74	14	78	13.2	6.26	3	33			
	EG12	2.6	62	34.5	12.84	15	66	35.6	13.23	19	66	12.1	5.76	4	30			
	AL894	2.36	1048	35.9	23.63	6	134	25.1	17.57	2	106	8.0	6.40	1	45			
	LA2C	2.34	500	38.0	15.00	12	96	35.0	14.00	7	128	11.0	5.00	3	28			
	Omo57	2.34	44	24.8	10.55	10	58	20.4	6.85	10	44	7.7	4.01	1	18			
	Omo123	2.34	110	20.8	7.50	7	50	17.8	6.49	6	38	5.9	2.79	1	16			
	DK	> 1.84	115	40.2	14.80	18	111	37.4	11.22	17	71	11.9	5.40	4	29			
	FLKZinj	1.76-1.84	125	36.8	12.13	16	82	32.9	11.59	4	76	11.5	5.45	4	36			
	SCNP	NA	31	33.5	15.80	14.8	70.98	26.5	12.42	6	52.4	13.2	7.46	1.6	33.5			
	Cores																	
	LOM3	3.3	83	167.0	23.40	132	260	147.8	23.10	90	210	108.8	21.80	61	170			
	OGS7	2.6	7	44.1	13.68	28	67	59.0	8.54	45	70	37.0	8.20	22	49			
	EG10	2.6	16	83.3	10.34	69	105	60.9	9.18	44	80	45.3	12.36	30	69			
	EG12	2.6	7	74.5	8.72	58	93	59.7	8.06	49	77	43.7	7.740	25	53			
	AL894	2.36	38	75.0	30.32	19.31	136.3	55.3	22.54	12.21	94.9	35.9	18.10	7.92	78.2			
	LA2C	2.34	70	66.0	18.00	39	123	52.0	14.00	32	95	32.0	12.00	12	78			

a, Dimension data for all technological categories identified in this study. b, Metric comparison of SCNP capuchin SoS percussion flakes and flaked hammer-stones with hominin Pliocene–Pleistocene flake and core dimensions. Data and table adapted from Harmand *et al.* (2015).

Evolution of *Hoxa11* regulation in vertebrates is linked to the pentadactyl state

Yacine Kherdjemi^{1,2}, Robert L. Lalonde³, Rushikesh Sheth¹, Annie Dumouchel¹, Gemma de Martino^{1†}, Kyriel M. Pineault⁴, Deneen M. Wellik⁴, H. Scott Stadler⁵, Marie-Andrée Akimenko³ & Marie Kmita^{1,2,6}

The fin-to-limb transition represents one of the major vertebrate morphological innovations associated with the transition from aquatic to terrestrial life and is an attractive model for gaining insights into the mechanisms of morphological diversity between species¹. One of the characteristic features of limbs is the presence of digits at their extremities. Although most tetrapods have limbs with five digits (pentadactyl limbs), palaeontological data indicate that digits emerged in lobed fins of early tetrapods, which were polydactylous². How the transition to pentadactyl limbs occurred remains unclear. Here we show that the mutually exclusive expression of the mouse genes *Hoxa11* and *Hoxa13*, which were previously proposed to be involved in the origin of the tetrapod limb^{1–6}, is required for the pentadactyl state. We further demonstrate that the exclusion of *Hoxa11* from the *Hoxa13* domain relies on an enhancer that drives antisense transcription at the *Hoxa11* locus after activation by HOXA13 and HOXD13. Finally, we show that the enhancer that drives antisense transcription of the mouse *Hoxa11* gene is absent in zebrafish, which, together with the largely overlapping expression of *hoxa11* and *hoxa13* genes reported in fish^{3–7}, suggests that this enhancer emerged in the course of the fin-to-limb transition. On the basis of the polydactyly that we observed after expression of *Hoxa11* in distal limbs, we propose that the evolution of *Hoxa11* regulation contributed to the transition from polydactyl limbs in stem-group tetrapods to pentadactyl limbs in extant tetrapods.

Several studies provided evidence for the implication of Hox genes in the fin-to-limb transition^{8–13}, notably *Hoxa13* and *Hoxd13* (*Hox13* hereafter), which are required for digit morphogenesis^{10–14}. Comparison of their expression pattern in fin and limb buds revealed a significant expansion of the *Hox13* domain in distal limbs¹⁵ and engineered enlargement of the *Hoxd13* domain in fish resulted in more chondrogenic tissue forming distally as well as fin fold reduction¹²—that is, morphological changes associated with the fin-to-limb transition. It was thus proposed that the evolution of *Hox13* regulation has likely been instrumental to the emergence of the limb characteristic feature, that is, the digits^{10,12}. In mice, this regulation relies on a series of remote transcriptional enhancers^{16,17}, and although a subset of these enhancers exists in fish¹⁸, the expansion of the *Hox13* domain in limb was probably associated with the emergence of tetrapod-specific enhancers during the fin-to-limb transition^{10–13}. Another notable difference is the mutually exclusive expression of *Hoxa11* and *Hoxa13* in tetrapod limbs, contrasting with their largely overlapping expression in fins^{3–7}. Two hypotheses have been put forward to explain how *Hoxa11* gets proximally restricted in tetrapod limbs. One hypothesis suggested a *Hoxa13*-dependent repression of *Hoxa11* in the presumptive autopod^{9,13,19}, whereas the second proposed that antisense transcription at the *Hoxa11* locus prevents expression of the gene distally^{20–22},

but the functional importance of the mutually exclusive expression of *Hoxa11* and *Hoxa13* in tetrapod limbs is unknown.

Previous chromatin conformation analyses revealed that, in distal limbs, 5' *HoxA* genes (that is, *Hoxa9* to *Hoxa13*) are grouped within a chromatin sub-topological domain (sub-TAD) interacting with sub-TADs containing distal limb enhancers¹⁷. Yet, although *Hoxa10* and *Hoxa13* are both expressed distally, *Hoxa11* expression is proximally restricted (Fig. 1a–c), suggesting that *Hoxa11* is part of the distal limb regulatory landscape, but a specific, yet unknown, mechanism prevents its expression distally^{13,19}. To test this possibility, we first took advantage of a mouse line in which the *Hoxa11* gene is replaced by a PGK-neomycin resistance cassette²³, which we used as a reporter transgene. We found neomycin expression in distal limbs (Fig. 1d), indicating that *Hoxa11* proximal restriction is linked to specific feature(s) of the gene itself. We next analysed the putative implication of antisense long non-coding RNAs previously identified at the *Hoxa11* locus^{20,21} and robustly expressed in the distal limb bud²⁰ (Fig. 1e). Among the distinct *Hoxa11* antisense transcripts (*Hoxa11as*, also known as *Hoxa11os*), two initiate upstream of the *Hoxa11* gene and are thus non-overlapping with *Hoxa11* (*Hoxa11as-a*; Fig. 1e) and the other two initiate within *Hoxa11* exon 1 (*Hoxa11as-b*; Fig. 1f). Notably, only *Hoxa11as-b* expression pattern is mutually exclusive with *Hoxa11* expression domain (Fig. 1f, compare with Fig. 1b). To test whether antisense transcription overlapping with *Hoxa11* exon 1 prevents *Hoxa11* expression distally, we took advantage of the *Hoxa11*^{eGFP} mutant line, which lacks *Hoxa11as-b* start sites as the enhanced green fluorescent protein (eGFP) coding sequence replaces most of *Hoxa11* exon 1 (ref. 24). This mutation disrupted antisense transcription normally initiating 3' to *Hoxa11* promoter (Extended Data Fig. 1a, b) while *gfp* expression driven by the *Hoxa11* promoter was present both in the proximal and distal domains (Fig. 1g). By contrast, ectopic expression of *Hoxa11as-b* in the entire limb had no effect on *Hoxa11* expression (Extended Data Fig. 2c–e), thereby excluding a *trans*-acting effect of *Hoxa11as-b* on *Hoxa11* expression. Together, our data suggest that *Hoxa11* distal repression is due to the antisense transcription event or the antisense *Hoxa11as-b* transcripts acting in *cis*.

Previous mapping of active enhancers in distal limbs¹⁷ (referred to as 'digit' enhancers hereafter) uncovered a putative 'digit' enhancer embedded in *Hoxa11* intron. We thus proposed that this enhancer might control *Hoxa11as-b* expression. We first tested the transcriptional enhancer activity of this DNA region in transgenic embryos and confirmed its ability to act as a transcriptional enhancer in distal limbs (Fig. 2a). Next, we generated mutant mice lacking this enhancer (*Hoxa11*^{ΔInt/ΔInt}; Extended Data Fig. 2) to examine its potential implication in *Hoxa11as-b* expression. Analysis of antisense transcription in *Hoxa11*^{ΔInt/ΔInt} limbs showed no detectable expression of *Hoxa11as-b* in the most distal cells (Fig. 2b, c), indicating that in these cells, the

¹Genetics and Development Research Unit, Institut de Recherches Cliniques de Montréal, Montréal, Québec H2W 1R7, Canada. ²Département de Médecine (Programme de Biologie Moléculaire), Université de Montréal, Montréal, Québec H3T 1J4, Canada. ³Department of Biology and CAREG, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. ⁴Department of Internal Medicine, Division of Molecular Medicine and Genetics, University of Michigan, Ann Arbor, Michigan 48109-2200, USA. ⁵Shriners Hospital for Children, Portland, Oregon 97239, USA. ⁶Department of Experimental Medicine, McGill University, Montreal, Quebec H3A 1A3, Canada. †Present address: Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada.

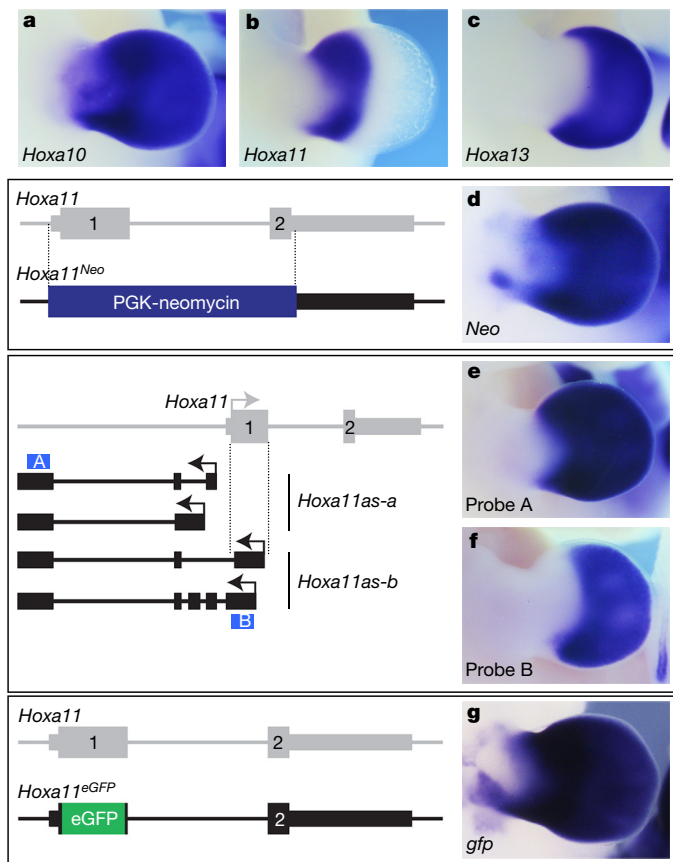


Figure 1 | The proximal restriction of *Hoxa11* is linked to antisense transcription at the *Hoxa11* locus. **a–c**, Expression of *Hoxa10* (**a**), *Hoxa11* (**b**) and *Hoxa13* (**c**) in wild-type limb bud from embryonic day (E) 11.5 mouse. **d**, Replacement of the *Hoxa11* gene with the PGK-neomycin cassette (*Hoxa11*^{Neo}; scheme to the left), results in neomycin expression both in the proximal and distal domains. **e**, **f**, Expression of all antisense transcripts (**e**) and antisense transcripts overlapping with *Hoxa11* exon 1 (**f**) in E11.5 wild-type limb. Schemes of the antisense transcripts and the probes used (blue boxes) are on the left. Note that the antisense transcripts overlapping with *Hoxa11* exon 1 (*Hoxa11as-b*) are distally restricted (**f**), reminiscent of *Hoxa13* expression (**c**) and mutually exclusive with the *Hoxa11* pattern (**b**). **g**, Deletion of the antisense transcript start sites in *Hoxa11* exon 1, via replacement of most of exon 1 with the eGFP coding sequence (*Hoxa11*^{eGFP}; scheme to the left) and expression of *gfp* under the control of the *Hoxa11* promoter (right). Original magnification, $\times 31.5$ (for all images).

identified enhancer is required for antisense transcription overlapping with *Hoxa11* exon 1. Some *Hoxa11as-b* expression remained in proximal cells of the presumptive handplate (presumptive carpal region; Fig. 2c), which suggests that additional *cis*-regulatory element(s) trigger antisense transcription in these cells. Notably, the deletion of the enhancer abrogating *Hoxa11as-b* expression in the most distal cells also resulted in ectopic expression of *Hoxa11* in the presumptive digits (Fig. 2d, e). The gain-of-sense transcription in *Hoxa11*^{eGFP} distal limbs (Fig. 1g) indicates that it is not the intronic regulatory region per se but *Hoxa11as-b* expression or the antisense transcription event that represses *Hoxa11* expression distally.

Analysis of the enhancer sequence revealed several putative binding sites for HOXA13, the expression of which occurs in digit progenitor cells²⁵ and is required in conjunction with HOXD13 for digit morphogenesis¹⁴. Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) indicated that, in distal limb cells, HOXA13 as well as HOXD13 bind to the identified enhancer (Extended Data Fig. 3a). Moreover, transcription assay in 293T cells shows that HOXA13 has a positive effect on the enhancer activity (Extended

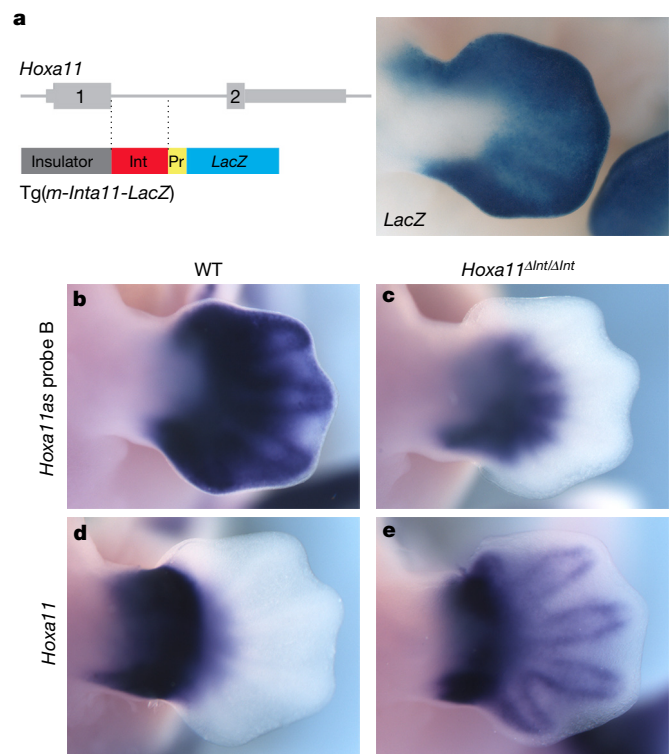


Figure 2 | Deletion of the distal enhancer in *Hoxa11* intron results in impaired antisense transcription and gain of sense transcription in distal cells. **a**, Left, scheme of the Tg(*m-Inta11-LacZ*) transgene carrying the predicted distal enhancer (Int, red box). Right, X-gal staining of E12.5 transgenic embryos ($n = 5$). **b–e**, Expression of *Hoxa11as-b* (**b**, **c**) and *Hoxa11* (**d**, **e**) in wild-type (WT; **b**, **d**) and *Hoxa11*^{ΔInt/ΔInt} (**c**, **e**) mouse limbs at E12.5. Note that based on the observed gain of *Hoxa11* expression, other regulatory input(s) could be implicated in *Hoxa11* regulation in distal cells. Pr, minimal promoter. Original magnification, $\times 31.5$ (for all images).

Data Fig. 3b). Together, these results raised the possibility that distal *Hoxa11* antisense transcription relies on HOX13. We thus analysed *Hoxa11* antisense transcription in the *Hoxa13*;*Hoxd13* allelic series. We used the probe recognizing all antisense transcripts such that expression in the proximal limb, where *Hox13* genes are not expressed, served as internal control. We found that although antisense transcription is barely modified in single mutants (Extended Data Fig. 4), it markedly decreases in the *Hoxa13*^{−/−} *Hoxd13*^{+/−} mutant (Fig. 3c, compare to Fig. 3a), and is completely abrogated in *Hoxa13*^{−/−} *Hoxd13*^{−/−} distal limbs (Fig. 3e). Analysis of the distal-specific antisense transcripts (*Hoxa11as-b*) confirmed that distal antisense transcription requires HOX13 function (Extended Data Fig. 5). Importantly, concomitant with the abrogation of antisense transcription, *Hoxa11* expression was gained distally (Fig. 3d–f, compare with Fig. 3b) consistent with the requirement of antisense transcription for *Hoxa11* proximal restriction.

To assess the functional significance of the HOXA13/D13-mediated repression of *Hoxa11*, we investigated the phenotypic outcome of distal *Hoxa11* expression. Although the deletion of the enhancer driving antisense transcription results in *Hoxa11* expression in distal limbs, the deletion extends up to the exon 1–intron boundary, thereby precluding the use of this mutant line to assess the phenotype resulting from distal *Hoxa11* expression. We thus generated a *Hoxa11* conditional gain-of-function allele (*Rosa26*^{*Hoxa11*}; Extended Data Fig. 6) to express *Hoxa11* ectopically and distally. We found that embryos carrying the *Rosa26*^{*Hoxa11*} allele and either *Hoxa13*;*Cre* (ref. 25) or *Prx1*;*Cre* (ref. 26) have limbs with extra digits (Fig. 3g, h), including postaxial extra digits (arrow in Fig. 3h and Extended Data Fig. 7). While some

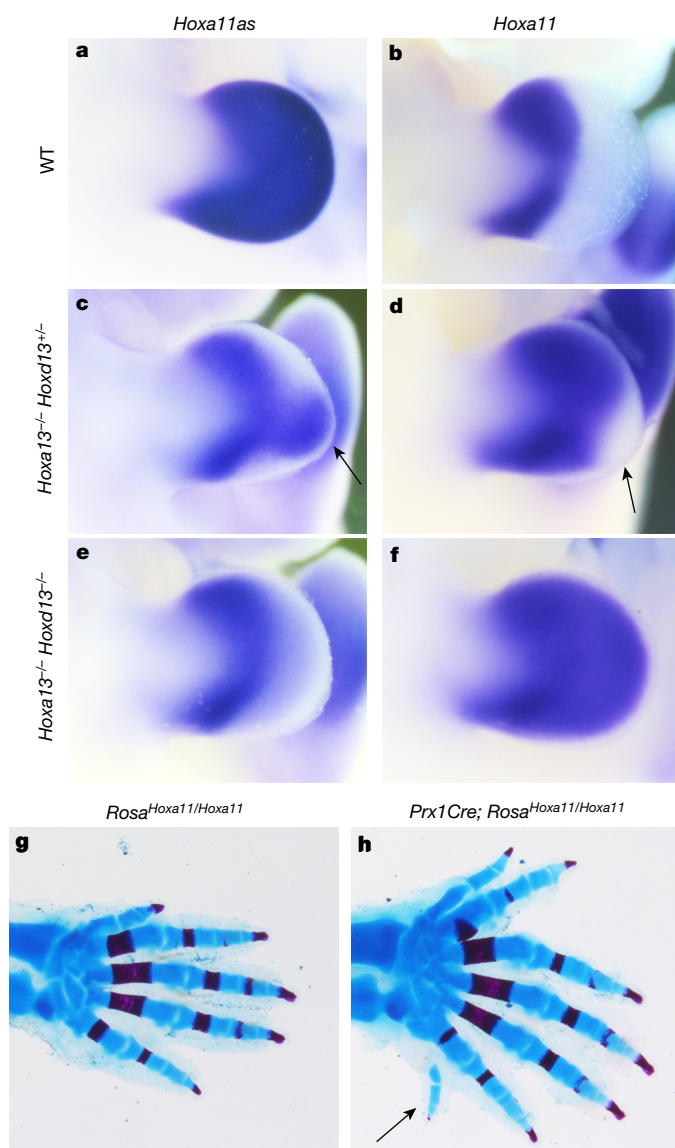


Figure 3 | *Hox13* inactivation disrupts *Hoxa11* antisense transcription in distal cells and distal *Hoxa11* expression results in the formation of supernumerary digits. **a–f**, *Hoxa11as* (probe A) (**a**, **c**, **e**) and *Hoxa11* (**b**, **d**, **f**) expression in E11.5 limb buds from wild-type (**a**, **b**), *Hoxa13*^{−/−} *Hoxd13*^{+/-} (**c**, **d**) and *Hoxa13*^{−/−} *Hoxd13*^{−/−} (**e**, **f**) mouse embryos. Arrows in **c** and **d** show the group of cells still expressing *Hoxa11as* in *Hoxa13*^{−/−} *Hoxd13*^{+/-} limbs (**c**), which corresponds to distal cells in which *Hoxa11* expression is not gained (**d**). **g**, **h**, Skeleton of *Rosa*^{*Hoxa11/Hoxa11*} (**g**) and *Prx1Cre*; *Rosa*^{*Hoxa11/Hoxa11*} (**h**) distal forelimb at postnatal day 0 (P0). Anterior is up. Original magnification, ×31.5 (**a–f**) and ×20 (**g**, **h**).

variations in the digit phenotype were observed among individuals, all homozygous mutants analysed were polydactylous (Extended Data Fig. 7c–e). Increased expression of *Hoxd11* in the presumptive autopod in the absence of *Hoxd13* also resulted in polydactyly, whereas a similar gain of *Hoxd10* or *Hoxd12* had no effect on digit number²⁷. These data raise the possibility that the formation of extra digits upon ectopic expression of *Hoxa11* or *Hoxd11* distally reflects the divergence between *Hoxa11/Hoxd11* targets and those of the other 5' *HoxA/D* genes. Notably, the evidence that *Hoxa11* expression in the distal limb results in the formation of extra digits indicates that the proximal restriction of *Hoxa11* expression is required for the pentadactyl state.

In contrast to the mutually exclusive *Hoxa11* and *Hoxa13* pattern in tetrapod limbs, *hoxa11* and *hoxa13* gene expression is largely

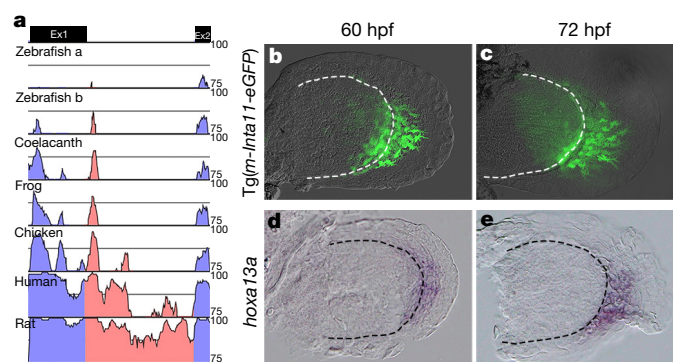


Figure 4 | The mouse *Hoxa11* antisense enhancer is functional in distal fins. **a**, mVISTA sequence conservation plot of the mouse *Hoxa11* intron (red) with tetrapod (rat, human, chicken and frog) and fish representatives (zebrafish **a**, **b**). **b**, **c**, GFP expression in fin buds of *Tg(m-Inta11-eGFP)* transgenic zebrafish embryos at 60 hpf (**b**) and 72 hpf (**c**), revealing the enhancer activity of the mouse *Hoxa11* intron in fish. Note the filopodia-like protrusions in GFP⁺ mesenchymal cells suggestive of a migration towards the fin fold. **d**, **e**, *hoxa13a* expression in developing fins at 60 hpf (**d**) and 72 hpf (**e**). Original magnification, ×400.

overlapping in zebrafish fins^{3–7} (Extended Data Fig. 8) as well as in other teleosts²⁸ (the medaka *Oryzias latipes*) and in fish models of both chondrichthyans⁵ (*Scyliorhinus canicula*) and basal actinopterygians³ (*Polyodon spathula*). The HOXA13/D13-mediated repression of *Hoxa11* identified in distal limb cells was thus probably implemented after the separation of actinopterygians and chondrichthyans, during the evolution of vertebrates towards tetrapod species. Consistent with this hypothesis, no *hoxa11* antisense transcription has been reported in fish^{22,29} (Extended Data Fig. 9). Moreover, sequence comparison of the mouse *Hoxa11* intron showed robust conservation among tetrapods, whereas considerably weaker sequence conservation was observed with fish *hoxa11* orthologues (Fig. 4a). To examine whether the lack of *hoxa11* antisense transcription in fish could be due to the absence of a distal enhancer within *hoxa11* intron, we tested the zebrafish *hoxa11a* and *hoxa11b* intronic sequences for potential enhancer activity using transgenic reporter assays in both zebrafish and mice. Neither the *hoxa11a* nor *hoxa11b* intron was capable of triggering expression of a reporter gene in fin nor in mouse limb buds (Extended Data Table 1), indicating that there is no distal enhancer in *hoxa11a* nor *hoxa11b* intron. By contrast, when we tested the transcriptional activity of the mouse *Hoxa11* intron in zebrafish, the analysis of four stable transgenic lines revealed that the mouse *Hoxa11* intron was able to drive reporter gene expression in the pectoral fin mesenchyme (Fig. 4b, c). At 60 hours post-fertilization (hpf), eGFP-positive cells were present at the distal rim of the endoskeletal disc and migrating into the fin fold (Fig. 4b) and by 72 hpf most eGFP-positive cells were found in the fin fold mesenchyme (Fig. 4c). The expression of the reporter transgene was reminiscent of *hoxa13a* expression at 60 hpf (Fig. 4d) and 72 hpf (Fig. 4e), indicating that the mouse enhancer in *Hoxa11* intron was active in the *Hoxa13* domain also in zebrafish. Together, our data indicate that all the transcription factors required for the activity of the mouse enhancer are present in zebrafish fins, and that the enhancer driving *Hoxa11* antisense transcription does not exist in the intron of the zebrafish *hoxa11a* and *hoxa11b* genes. We therefore propose that the emergence of the enhancer triggering *Hoxa11* antisense transcription, and thus distal repression of *Hoxa11*, occurred in the course of evolution towards tetrapod species.

In summary, our work reveals that the mutually exclusive expression of *Hoxa11* and *Hoxa13* in tetrapods is associated with the emergence of a transcriptional enhancer in *Hoxa11* intron, which upon HOXA13/D13-dependent activation, triggers antisense transcription and thereby

prevents *Hoxa11* expression distally. On the basis of the evidence that this HOX13-mediated regulation of *Hoxa11* probably emerged during the fin-to-limb transition and the polydactyly resulting from distal expression of *Hoxa11* in mice, we propose that the evolution of *Hoxa11* regulation has contributed to the transition from polydactyly in stem-group (extinct) tetrapods to pentadactyly in extant tetrapods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 April 2014; accepted 26 August 2016.

Published online 5 October 2016.

- Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
- Coates, M. I., Jeffery, J. E. & Rut, M. Fins to limbs: what the fossils say. *Evol. Dev.* **4**, 390–401 (2002).
- Davis, M. C., Dahn, R. D. & Shubin, N. H. An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. *Nature* **447**, 473–476 (2007).
- Metscher, B. D. *et al.* Expression of Hoxa-11 and Hoxa-13 in the pectoral fin of a basal ray-finned fish, *Polyodon spathula*: implications for the origin of tetrapod limbs. *Evol. Dev.* **7**, 186–195 (2005).
- Sakamoto, K. *et al.* Heterochronic shift in Hox-mediated activation of sonic hedgehog leads to morphological changes during fin development. *PLoS One* **4**, e5121 (2009).
- Sordino, P., Duboule, D. & Kondo, T. Zebrafish *Hoxa* and *Evx-2* genes: cloning, developmental expression and implications for the functional evolution of posterior *Hox* genes. *Mech. Dev.* **59**, 165–175 (1996).
- Ahn, D. & Ho, R. K. Tri-phasic expression of posterior *Hox* genes during development of pectoral fins in zebrafish: implications for the evolution of vertebrate paired appendages. *Dev. Biol.* **322**, 220–233 (2008).
- Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
- Wagner, G. P. & Chiu, C. H. The tetrapod limb: a hypothesis on its origin. *J. Exp. Zool.* **291**, 226–240 (2001).
- Sordino, P., van der Hoeven, F. & Duboule, D. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).
- Freitas, R., Zhang, G. & Cohn, M. J. Biphasic *Hoxd* gene expression in shark paired fins reveals an ancient origin of the distal limb domain. *PLoS One* **2**, e754 (2007).
- Freitas, R., Gómez-Marín, C., Wilson, J. M., Casares, F. & Gómez-Skarmeta, J. L. Hoxd13 contribution to the evolution of vertebrate appendages. *Dev. Cell* **23**, 1219–1229 (2012).
- Woltering, J. M., Noordermeer, D., Leleu, M. & Duboule, D. Conservation and divergence of regulatory strategies at Hox loci and the origin of tetrapod digits. *PLoS Biol.* **12**, e1001773 (2014).
- Fromental-Ramain, C. *et al.* Hoxa-13 and Hoxd-13 play a crucial role in the patterning of the limb autopod. *Development* **122**, 2997–3011 (1996).
- Schneider, I. & Shubin, N. H. The origin of the tetrapod limb: from expeditions to enhancers. *Trends Genet.* **29**, 419–426 (2013).
- Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145 (2011).
- Berlivet, S. *et al.* Clustering of tissue-specific sub-TADs accompanies the regulation of *HoxA* genes in developing limbs. *PLoS Genet.* **9**, e1004018 (2013).
- Gehrke, A. R. *et al.* Deep conservation of wrist and digit enhancers in fish. *Proc. Natl Acad. Sci. USA* **112**, 803–808 (2015).
- Sheth, R., Bastida, M. F., Kmita, M. & Ros, M. “Self-regulation,” a new facet of *Hox* genes’ function. *Dev. Dyn.* **243**, 182–191 (2014).
- Hsieh-Li, H. M. *et al.* Hoxa 11 structure, extensive antisense transcription, and function in male and female fertility. *Development* **121**, 1373–1385 (1995).
- Potter, S. S. & Branford, W. W. Evolutionary conservation and tissue-specific processing of Hoxa 11 antisense transcripts. *Mamm. Genome* **9**, 799–806 (1998).
- Leite-Castro, J., Bevilacqua, V., Rodrigues, P. N., Freitas, R. & Hox, A. Genes and the fin-to-limb transition in vertebrates. *J. Dev. Biol.* **4**, 10 (2016).
- Small, K. M. & Potter, S. S. Homeotic transformations and limb defects in *Hox A11* mutant mice. *Genes Dev.* **7**, 2318–2328 (1993).
- Nelson, L. T., Rakshit, S., Sun, H. & Wellik, D. M. Generation and expression of a *Hoxa11* *Cre* targeted allele in mice. *Dev. Dyn.* **237**, 3410–3416 (2008).
- Scotti, M., Kherdjemil, Y., Roux, M. & Kmita, M. A *Hoxa13:Cre* mouse strain for conditional gene manipulation in developing limb, hindgut, and urogenital system. *Genesis* **53**, 366–376 (2015).
- Logan, M. *et al.* Expression of *Cre* recombinase in the developing mouse limb bud driven by a *Prl* enhancer. *Genesis* **33**, 77–80 (2002).
- Kmita, M., Fraudeau, N., Héroult, Y. & Duboule, D. Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. *Nature* **420**, 145–150 (2002).
- Takamatsu, N. *et al.* Duplicated *Abd-B* class genes in medaka *hoxAa* and *hoxAb* clusters exhibit differential expression patterns in pectoral fin buds. *Dev. Genes Evol.* **217**, 263–273 (2007).
- Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–1550 (2011).

Acknowledgements We thank Q. Zhu and L. Lian from the IRCM transgenic core facility for the ES cell work and production of transgenic mouse lines. We are particularly grateful to A. Kania for critical reading of the manuscript as well as laboratory members for insightful discussions and sharing reagents. This work was supported by the Canadian Institute for Health Research (MOP-115127) and the Canada Research Chair program to M.K. (RCHS0192), the Natural Sciences and Engineering Research Council of Canada (155817-2012) to M.-A.A. and Shriners Hospital Research grant 85400 to H.S.S. Y.K. was supported by a fellowship from the Molecular Biology program of the Université de Montréal and the IRCM fellowship Michel-Bélanger. R.S. was supported by a post-doctoral fellowship from the Canadian Institute for Health Research. D.M.W. and K.M.P. were supported by NIH NIAMS AR061402, with K.M.P. additionally supported by NIH T32 DE007057.

Author Contributions Y.K. and M.K. conceived the study and analysed the data. Y.K. designed and conducted all mouse experiments with the help of R.S. for the generation of the mouse lines. All fish experiments were performed by R.L.L. under the supervision of M.-A.A. R.S. performed the ChIP-seq experiments. A.D. provided technical help for the mouse experiments. G.M. performed preliminary experiments related to Figs 2a and 3c, e. D.M.W. and K.M.P. provided *Hoxa11*^{eGFP/eGFP} embryos. H.S.S. provided the HOXA13 and HOXD13 antibodies. M.K. wrote the paper. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. (marie.kmita@ircm.qc.ca).

Reviewer Information *Nature* thanks R. Freitas, J. L. Gomez-Skarmeta and S. Mackem for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size.

Mouse lines. *Hoxa11*^{Neo}, *Hoxa11*^{eGFP}, *Hoxa13*^{null} (*Hoxa13*^{Str}) and *Hoxd13*^{null} (*Hoxd13*^{lacZ}) mouse lines were previously described^{14,23,24,30}.

Rosa^{*Hoxa11*} knock-in allele was constructed as follows: PacI-AscI fragment from pBTG (Addgene plasmid 15037)³¹ was inserted into the previously described *Rosa26* targeting vector³² pROSA26Am1 (Addgene plasmid 15036)³¹. The mouse *Hoxa11* cDNA was inserted at the SmaI site within the MCS. The vector was linearized by SmaI digest prior electroporation into embryonic stem (ES) cells. After double selection using G418 and DTA negative selection, 96 ES cell clones were analysed by Southern blot for homologous recombination. Two independent clones were injected into blastocysts obtained from C57BL/6J mice, subsequently implanted into pseudo-pregnant females. After germline transmission of the *Rosa*^{*Hoxa11*} allele, mice and embryos were genotyped by Southern blot (a scheme with restriction sites and probes used is presented in Extended Data Fig. 6) and PCR. The following PCR primers were used: fw_wt: 5'-GCAATACCTTTCTGGGAGTTCT-3', rev_wt: 5'-TCGGGTGAGCATGTCTTTTAATC-3', rev_flox: 5'-TTCAATGGCCGATCCCATATT-3', rev_del: 5'-AGGTTGGAGGAGTAGGAGTATG-3'. Wild-type band: 384 bp, flox band: 881 bp, del band: 583 bp. The moderate transcription resulting from the *Rosa26* promoter allowed for *Rosa26*^{*Hoxa11*} expression at a level comparable to the *Hoxa11* gain observed in our series of mutants.

Hoxa11^{Δint} mouse line was generated through pronuclear injection of single-guide RNAs (sgRNAs). We used the CRISPR (http://crispr.mit.edu/) platform to design sgRNAs flanking the region to delete. Complementary strands were annealed, phosphorylated and cloned into the BbsI site of pX330 CRISPR/Cas9 vector (Addgene plasmid 42230)³³. SgInt1_fw: 5'-CACCGACTCCCTTTTCATAAGCCC-3'; SgInt1_rev: 5'-AAACGCGCTTTATGAAAGGGAGTC-3'; SgInt2_fw: 5'-CACCGAGCAACAGGCGAGTTTGC-3'; SgInt2_rev: 5'-AAACGCGCAAACTCGCTGTTGTC-3'. Mice and embryos were genotyped by Southern blot (a scheme with restriction sites and probe used is presented in Extended Data Fig. 2) as well as PCR. The Southern blot probe corresponds to the ScaI-HpaI fragment in the 3' untranslated region (UTR) of the *Hoxa11* gene. Primers used for PCR genotyping, fw: 5'-GGCCACCTAAGGAAGGAGAG-3'; rev: 5'-GGCTCCGGTGCCTATAAG-3'.

Three *Prx1-Hoxa11* transgenic lines were derived from three distinct founders obtained from pronuclear injection of the *Prx1-Hoxa11* transgene. The *Prx1-Hoxa11* transgene carries the *Prx1* promoter upstream of the mouse *Hoxa11* (GenBank: U20367.1 and U20366.1) and the SV40 polyadenylation sequence was inserted downstream *Hoxa11*. Embryos were genotyped by PCR using DNA from the amniotic membrane and the following pair of primers: fw: 5'-CTTCTCTCTGGCTCTGATG-3' and rev: 5'-GACAAGAAGCCGAGAA-3' (for U20367.1) or fw: 5'-GTCCGAGGAAAGGAGGTAG-3' and rev: 5'-GCTCCTCTAACATGTATTG-3' (for U20366.1).

All mice were of mixed background (C57BL/6 X 129).

The Tg(*m-Inta11-LacZ*) transgene was generated by subcloning the mouse *Hoxa11* intron upstream of the *Hbb* (β-globin) minimal promoter and a LacZΔCpG NLS reporter. The *H19* insulator was inserted upstream of the *Hoxa11* intron. Tg(*m-Inta11-LacZ*) embryos were produced by pronuclear injection.

Whole-mount *in situ* hybridization, X-gal staining, skeletal preparations and imaging. For skeletal preparation, newborn mice were processed using the standard alcian blue alizarin red staining protocol³⁴ ($n = 10$ for each genotype).

Whole-mount *in situ* hybridizations were performed using previously described protocol³⁵ and probes³⁵ (*gfp*³⁶, *Neo*, *Hoxa11*, *Hoxa13*). Embryos were genotyped prior *in situ* hybridization (no blinding). *Hoxa11* probes were generated using limb cDNA and the following primers: fw 5'-AGAGGCGCTGAGGAGCCTTCTC-3' and rev 5'-GGCCGCTGTGGACACTAGCATATACC-3' (probe A); fw 5'-CCTTCTCGGCGTTCTGTGC-3' and rev 5'-GGCATACTCTACTCTCAACCT-3' (probe B).

X-gal staining was performed using standard protocol³⁵. Embryos were genotyped after X-gal staining (which results in blinding test).

All mouse specimens were imaged using the Leica DFC450C camera. For each experiment, a minimum of three embryos per genotype was used as we considered that reproducible staining/expression patterns with three distinct embryos of the same genotype are significant. The experiments shown were repeated at least twice. We did not use the randomization method.

Subcloning of zebrafish *hoxa11a/b* intron and microinjections in zebrafish embryos. The zebrafish *hoxa11a* (713 bp; gene ID 58061, NCBI) and *hoxa11b* (747 bp; gene ID 30382, NCBI) introns were amplified from zebrafish genomic DNA using the following primers: *hoxa11a* intron: fw 5'-GAATTCAACAGTAAG

TACGAGCTCAAC-3'; rev 5'-GGTACCACCTAAATGTAAATACACGT-3'; *hoxa11b* intron: fw 5'-GAATTCCAGCGCAGCAGCAGTACGT-3'; rev 5'-GGTACCCCGTGTCTTTTGCCATCTAA-3'.

The zebrafish *hoxa11a* and *hoxa11b* and the mouse *Hoxa11* introns were subcloned into the pEGFP-N1 vector (CLONTECH Laboratories, Inc.) in which the CMV promoter upstream of eGFP was replaced with the human *HBB* minimal promoter using the following primers: fw 5'-GGATCCCTGGGCATAAAAGTCAG-3', rev 5'-ACCGGTTCTGCTTCTGGAAGGCT-3'. This vector also contains the Tol2 arms to increase transgenesis efficiency. For screening purposes, a heart marker (*cmcl2:mCherry*³⁷) was added to zebrafish Tg(*z-Inta11a-eGFP*) and Tg(*z-Inta11b-eGFP*) constructs. All constructs were microinjected in one-cell stage wild-type zebrafish embryos at a concentration of 100 ng μl⁻¹ together with 50 ng μl⁻¹ transposase mRNA.

Generation of zebrafish transgenic lines. Primary injected zebrafish (P₁) are raised until 3 months of age, and then are screened for transgenic progeny (F₁). P₁ fish are crossed with wild-type fish and the embryos are screened at 2 days post-fertilization (dpf). Owing to lack of fin fold eGFP expression in the Tg(*z-Inta11a-eGFP*; *cmcl2:mCherry*), Tg(*z-Inta11b-eGFP*; *cmcl2:mCherry*) injected fish, embryos were screened for the presence of the *cmcl2:mCherry* heart marker and genotyped to confirm the presence of the *hoxa11a/b* intron:eGFP elements. The following primers were used for genotyping: *hoxa11a*: fw 5'-GGTACCACCTAAATGTAAATACACGT-3', rev (eGFP) 5'-GTCCTCCTTGAAGTCGATGC-3'; *hoxa11b*: fw 5'-GGTACCCCGTGTCTTTTGCCATCTAA-3', rev (eGFP) 5'-GTCCTCCTTGAAGTCGATGC-3'.

Three transgenic lines for Tg(*m-Inta11-eGFP*) were obtained to confirm the expression pattern. A fourth line containing the *cmcl2:mCherry* heart marker was also created. To confirm the *Hbb* minimal promoter does not drive tissue-specific expression alone, a transgenic line Tg(*HBB:eGFP*; *cmcl2:mCherry*) was also created and genotyped using the following primers: *Hbb*: fw 5'-GGATCCCTGGGCATAAAAGTCAG-3', rev (eGFP) 5'-GTCCTCCTTGAAGTCGATGC-3'.

Zebrafish *in situ* hybridization. *In situ* hybridization on whole-mount embryos was performed as previously described³⁸. Digoxigenin-labelled antisense RNA probes were generated using the following cDNAs: *hoxa13a* (500 bp; Addgene 36463), *hoxa13b* (700 bp; Addgene 36568), *hoxa11b* (probe 1) (Extended Data Fig. 8c, d); 800 bp; Addgene 36466). For *hoxa11a/b* antisense/sense RNA probes (Extended Data Fig. 9a, b), *hoxa11a* (713 bp; Gene ID 58061, NCBI) and *hoxa11b* (747 bp; gene ID 30382, NCBI) partial cDNAs (exon 1) were obtained by PCR with reverse transcription from total RNA of 24–48 hpf embryos using the following primers: *hoxa11a* exon 1: fw 5'-ATGATGGATTTTGACGAAAGGGTT-3', rev 5'-TGTTCCACCGCTAGTTTTTTCCT-3'; *hoxa11b* exon 1: fw 5'-ATGATGGATTTTGATGAGCGGGTA-3', rev 5'-TGCTGCTGCCGCTGAATTTATCTT-3'.

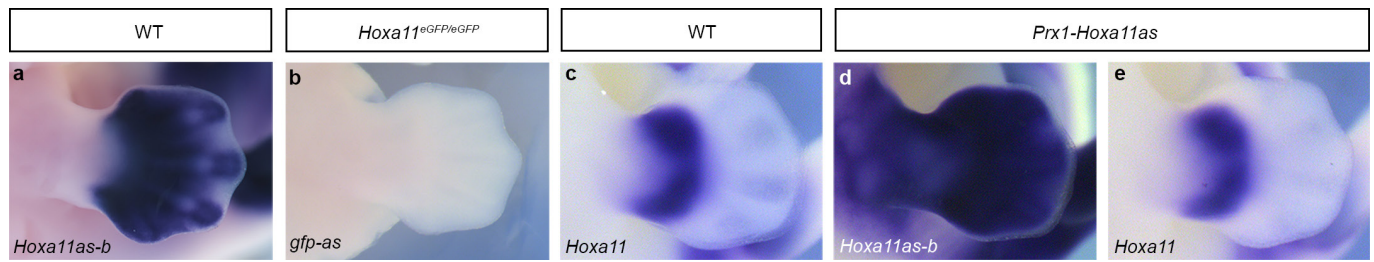
For accurate comparison, *hoxa11a* and *hoxa11b* sense and antisense probes, respectively, are identical in length and were transcribed using the same RNA polymerase. *In situ* hybridizations were also performed in parallel with identical staining times.

Transfection and gene expression analysis. 293T cells (ATCC) were transfected using lipofectamine. Cells (800,000) were plated in 6-well plates. Cells were checked for mycoplasma contamination using Venor GeM Mycoplasma Detection Kit (MP0025 SIGMA). A total of 2 μg of DNA (250 ng reporter plasmid, 250 ng effector plasmid or empty expression vector), 25 ng of mCherry expression vector as internal control and 1.45 μg carrier pBSK plasmid was used for each transfection. All transfections were performed in duplicates. Then, 24 h after transfection, the medium was changed and 48 h after transfection, cells were processed for RNA extraction. Reporter gene expression was normalized to internal control mCherry ($n = 3$). Gene expression (*Hoxa11*) was measured in dissected E11.5 forelimb buds of the *Rosa*^{*Hoxa11*} knock-in embryos that were stored in RNA later before RNA extraction ($n = 4$).

RNA extraction was done using RNeasy Plus mini kit (Qiagen 74134). cDNA synthesis was performed using M-MuLV reverse transcriptase (NEB) and a mix of random primers and oligo-dT on 1 μg of total RNA. Quantitative real-time-PCR was performed with cDNA and the SYBR Green kit (applied biosystems) using the following primers: fw 5'-AGGAGAAGGAGCGACGG-3' and rev 5'-GGTATTTTGGTATAAGGGCAGCG-3' (*Hoxa11*); fw 5'-CTTTGTCAAGCTCATTTCTTGG-3' and rev 5'-TCTTGCTCAGTGTCTTCTG-3' (*Gapdh*); fw 5'-TTGACCTAAAGACCATTCGACTTC-3' and rev 5'-TCTTCATGATGACTGCAGCAAA-3' (*Tbp*); fw 5'-GCCTACAACGTCACATCAAG-3' and rev 5'-GCGTTCGTACTGTTCAC-3' (mCherry); fw 5'-GACCCTGAAGTTTATCTGCA-3' and rev 5'-CCGTCGTCTTGAAGAAGA-3' (*gfp*).

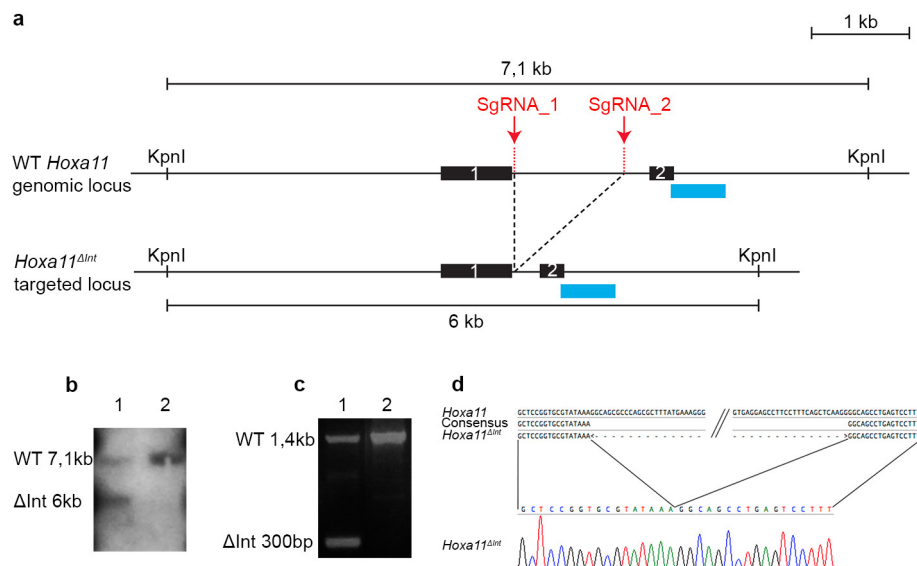
Study approval. All mice experiments described in this article were approved by the Animal Care Committee of the Institut de Recherches Cliniques de Montréal (protocols 2011-39 and 2014-14) and zebrafish experiments were approved by uOttawa Animal Care Committee (protocol BL-2317-R1).

30. Kmita, M., Kondo, T. & Duboule, D. Targeted inversion of a polar silencer within the HoxD complex re-allocates domains of enhancer sharing. *Nat. Genet.* **26**, 451–454 (2000).
31. Murtaugh, L. C., Stanger, B. Z., Kwan, K. M. & Melton, D. A. Notch signaling controls multiple steps of pancreatic differentiation. *Proc. Natl Acad. Sci. USA* **100**, 14920–14925 (2003).
32. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nat. Genet.* **21**, 70–71 (1999).
33. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
34. Sheth, R. *et al.* *Hox* genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science* **338**, 1476–1480 (2012).
35. Scotti, M. & Kmita, M. Recruitment of 5' *Hoxa* genes in the allantois is essential for proper extra-embryonic function in placental mammals. *Development* **139**, 731–739 (2012).
36. Mattar, P. *et al.* Basic helix-loop-helix transcription factors cooperate to specify a cortical projection neuron identity. *Mol. Cell. Biol.* **28**, 1456–1469 (2008).
37. Yelon, D., Horne, S. A. & Stainier, D. Y. Restricted expression of cardiac myosin genes reveals regulated aspects of heart tube assembly in zebrafish. *Dev. Biol.* **214**, 23–37 (1999).
38. Thisse, C. & Thisse, B. High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protocols* **3**, 59–69 (2008).



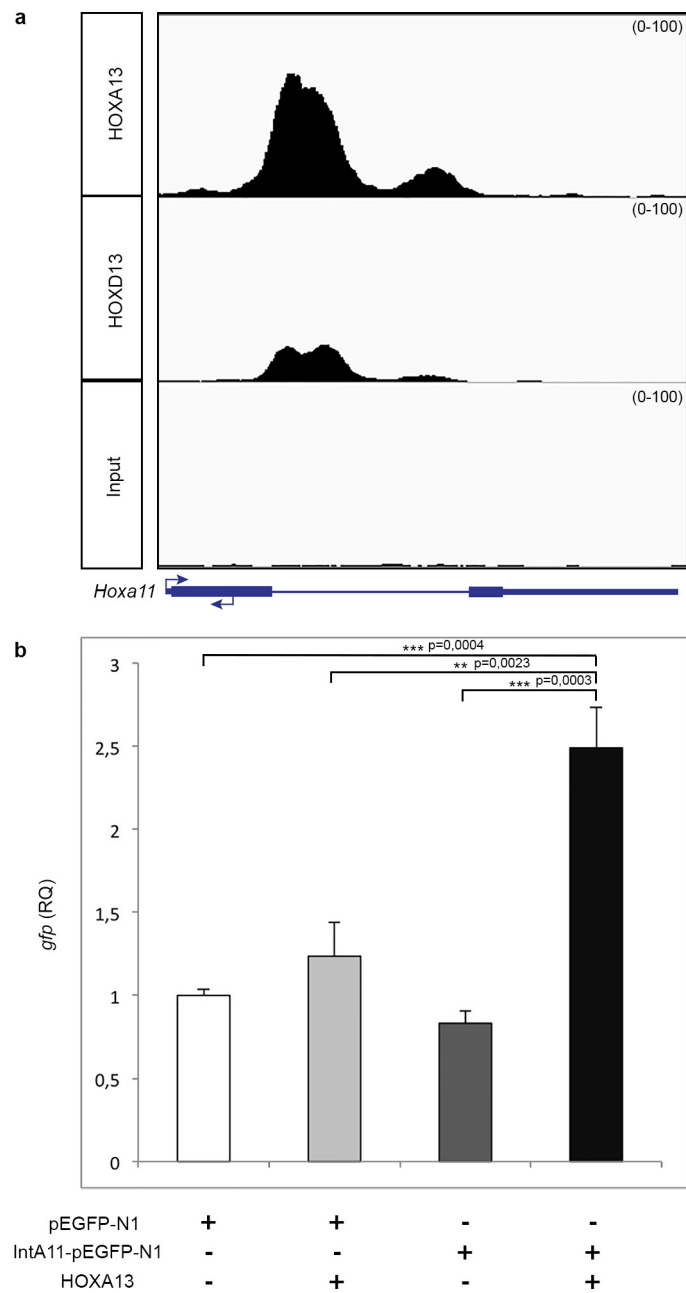
Extended Data Figure 1 | Absence of antisense transcription 3' to the *Hoxa11* promoter in the *Hoxa11*^{eGFP/eGFP} limb and evidence that *Hoxa11as-b* transcripts produced in *trans* have no effect on *Hoxa11* expression. **a, b**, Detection of *Hoxa11as-b* transcripts in wild-type limb buds at E12.5 (**a**), and whole-mount *in situ* hybridization to detect

gfp antisense transcripts in *Hoxa11*^{eGFP/eGFP} limb buds at E12.5 (**b**). **c–e**, *Hoxa11* expression in wild-type limb buds (**c**), and *Hoxa11as-b* (**d**) and *Hoxa11* (**e**) expression in *Prx1-Hoxa11as* limb buds. Original magnification, $\times 31.5$.

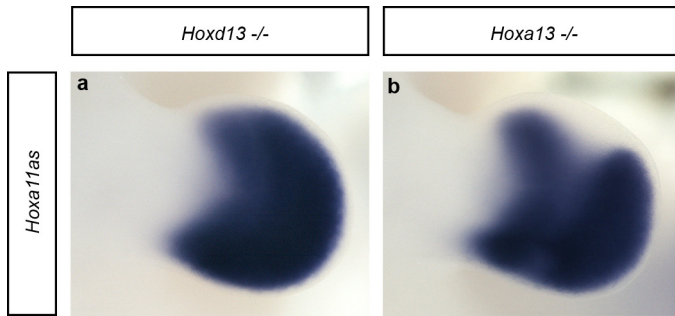


Extended Data Figure 2 | Deletion of the distal enhancer in *Hoxa11* intron using CRISPR-Cas9. **a**, Scheme of the wild-type and targeted (*Hoxa11*^{Δ_{int}}) loci. Sites targeted by the single-guide RNAs (sgRNA_1 and sgRNA_2) for the CRISPR-Cas9-mediated deletion of the distal enhancer. The blue rectangles indicate the position of the DNA probe used to confirm the deletion by Southern blot in **b**. **b**, Lane 1 shows the 6-kb KpnI band resulting from the CRISPR-Cas9-mediated deletion. Lane 2

was loaded with wild-type DNA. **c**, PCR reaction using a forward primer located upstream of sgRNA_1 and a reverse primer located downstream sgRNA_2 shows the presence of a 300 bp (Δ Int 300 bp) fragment expected for the *Hoxa11* ^{Δ Int} allele. **d**, The sequence of the 300-bp PCR fragment confirms the CRISPR-Cas9-mediated deletion of the *Hoxa11* intronic region containing the distal enhancer (only the sequence encompassing the deletion breakpoints is shown).

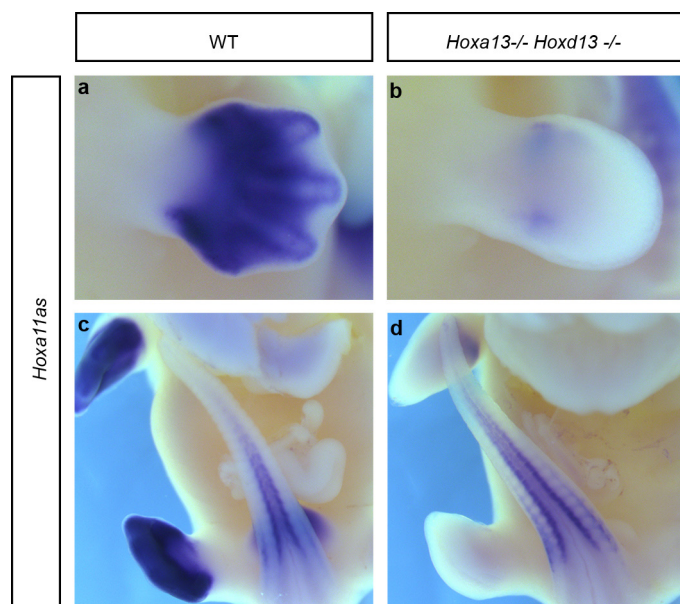


Extended Data Figure 3 | The distal enhancer located in the *Hoxa11* intron is bound by HOXA13 and HOXD13 in distal limb cells and its activity is increased by HOXA13 in 293T cells. **a**, Integrative genomics viewer (IGV) screenshot showing HOXA13 and HOXD13 ChIP-seq data at the *Hoxa11* locus. These ChIP-seq data were obtained using chromatin from distal forelimb buds of wild-type E11.5 mouse embryos (R. Sheth *et al.*, manuscript submitted). **b**, Transfection assay shows HOXA13 dependent activation of *Hoxa11* intron driving reporter gene expression. Two-tailed Tukey's multiple comparisons test was performed. Error bars indicate s.d ($n = 3$). RQ, relative quantification.

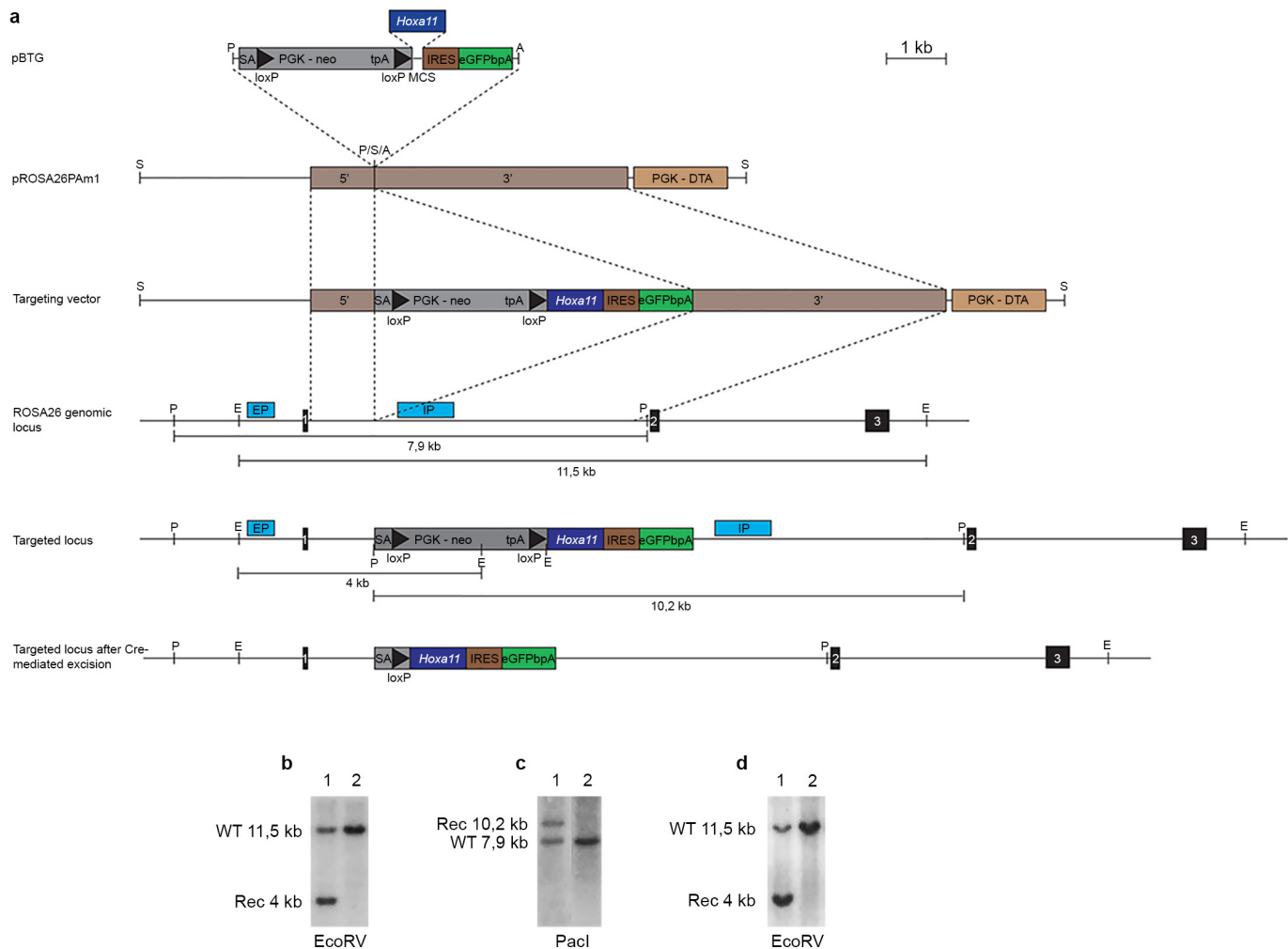


Extended Data Figure 4 | Individual inactivation of *Hoxa13* or *Hoxd13* is not sufficient to fully abrogate antisense transcription in distal limbs.

a, b, Whole-mount *in situ* hybridization, using probe A (see Fig. 1) to detect all antisense transcripts, on *Hoxd13*^{-/-} (**a**) and *Hoxa13*^{-/-} (**b**) mouse limb buds at E11.5. Antisense transcription in distal limbs remains robust in both mutants but a clear reduction is seen in the distal *Hoxa13*^{-/-} limbs. Original magnification, $\times 31.5$.

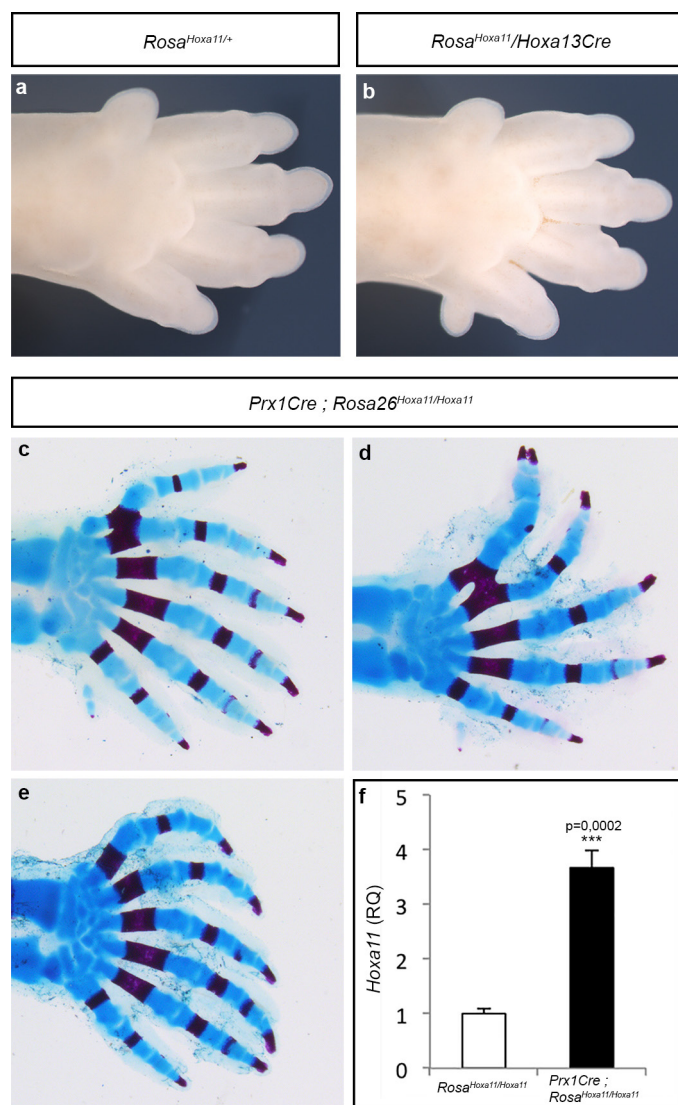


Extended Data Figure 5 | Inactivation of both *Hoxa13* and *Hoxd13* disrupts antisense transcription overlapping with the *Hoxa11* exon 1. **a–d**, *Hoxa11as-b* expression (probe B in Fig. 1) in limb buds (**a**, **b**) and tail buds (**c**, **d**) from wild-type (**a**, **c**) and *Hoxa13*^{-/-} *Hoxd13*^{-/-} (**b**, **d**) E12.5 mouse embryos. Whole-mount *in situ* hybridization shows that *Hoxa11as-b* expression in tail buds (internal control) is similar in both the wild-type (**c**) and double-mutant (**d**) embryos, whereas there is almost no expression remaining in *Hoxa13*^{-/-} *Hoxd13*^{-/-} limb buds (**b**). Original magnification, $\times 31.5$.



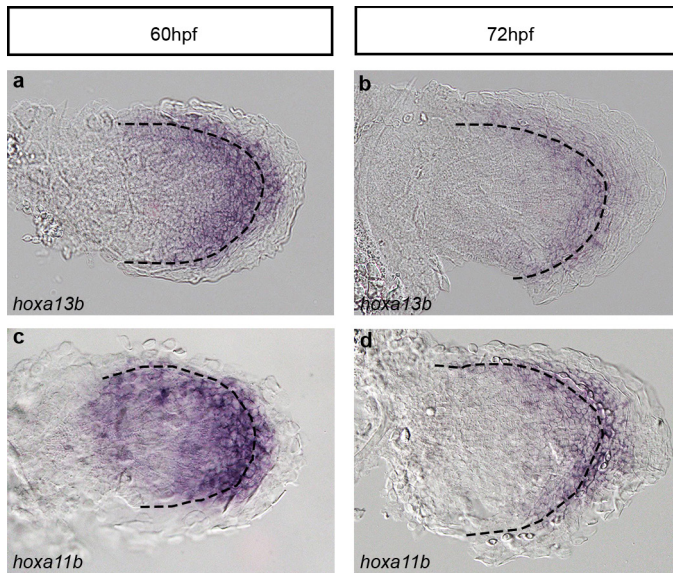
Extended Data Figure 6 | Generation of the *Rosa^{Hoxa11}* knock-in mouse line. **a, Targeting of the endogenous *Rosa26* locus (top three lines). The wild-type *Rosa26* locus is shown below (middle). Regions used as homologous arms for the recombination in ES cells are indicated by brown rectangles labelled 5' and 3', respectively. Scheme of the targeted locus after homologous recombination in ES cells and after Cre-mediated recombination is shown at the bottom. The position of the internal (IP)**

and external (EP) probes and restriction sites used for Southern blot analysis are indicated on both the wild-type and targeted locus. **b, c**, Southern blots of ES cells clones using the internal probe (**b**) and external probe (**c**) to detect the targeted allele (lane 1). **d**, Southern blot of wild-type (lane 2) and heterozygous (lane 1) mice. A, AscI; E, EcoRV; P, PacI; S, SwaI.



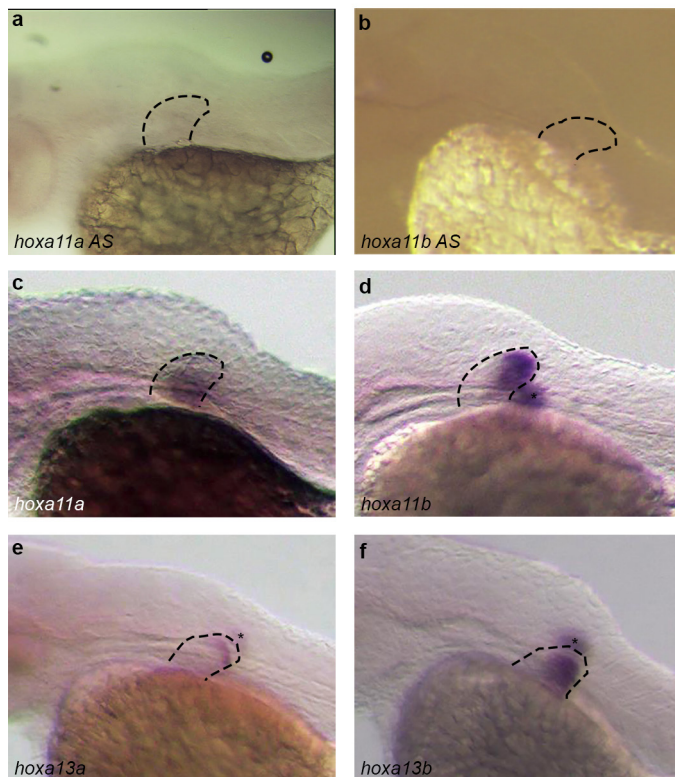
Extended Data Figure 7 | The conditional gain of *Hoxa11* using the *Hoxa13Cre* allele results in the formation of supernumerary digits.

a, b, Autopod of *Rosa^{Hoxa11/+}* (a) and *Rosa^{Hoxa11}/Hoxa13Cre* (b) at E15.5. Anterior is up. The *Rosa26* locus and *Hoxa13Cre* allele being on the same chromosome (Chr6), the gain-of-function phenotype was assessed with only one copy of the *Rosa^{Hoxa11}* allele. **c–e,** Autopod skeletons of *Prx1Cre; Rosa^{Hoxa11/Hoxa11}* mice at P0 from four distinct mutants (anterior is up). The number of digits varies from 6 to 7, with often a small post-axial extra-digit (posterior). The extra-digit phenotype is fully penetrant upon Cre-activation of two copies of the *Rosa^{Hoxa11}* allele ($n=10$). Original magnification, $\times 20$. **d,** Quantification of *Hoxa11* expression level by quantitative reverse transcriptase PCR (RT-qPCR) on RNA extracted from E11.5 forelimb, relative to both *Gapdh* and *Tbp* mRNA of *Prx1Cre; Rosa^{Hoxa11/Hoxa11}* embryos. Two-tailed *t*-test was performed. Error bars indicate s.d ($n=4$).



Extended Data Figure 8 | *hoxa11* and *hoxa13* are expressed in overlapping domains in zebrafish fins. a–d, Expression of *hoxa13b* (a, b) and *hoxa11b* (c, d) in zebrafish fins at 60 hpf (a, c) and 72 hpf (b, d). Dotted lines indicate the boundary between the endochondral disc and the fin fold. Original magnification, $\times 400$.

48hpf



Extended Data Figure 9 | Absence of antisense transcription at the *hoxa11a* and *hoxa11b* loci in zebrafish fins. **a, b,** Whole-mount *in situ* hybridization with probes designed to detect putative antisense transcription at *hoxa11a* (**a**) and *hoxa11b* (**b**). **c–f,** No antisense transcription is detected, whereas expression of *hoxa11a* (**c**), *hoxa11b* (**d**), *hoxa13a* (**e**) and *hoxa13b* (**f**) is observed in zebrafish fins at the same stage. Asterisks correspond to the staining from the fin on the other side of the embryo. Original magnification, $\times 63$.

Extended Data Table 1 | Summary of transient transgenic embryos analysed

Zebrafish Transient Transgenics	
Construct	% of eGFP positive fish
Tg(<i>HBB:eGFP</i>)	0% (n=74)
Tg(<i>z-Inta11a-eGFP</i>)	0% (n=105)
Tg(<i>z-Inta11b-eGFP</i>)	1.19% (n=84)
Tg(<i>m-Inta11-eGFP</i>)	91.9% (n=123)
Tg(<i>HBB:eGFP; cmc2:mCherry</i>)	1.25% (n=94)
Tg(<i>z-Inta11a-eGFP; cmc2:mCherry</i>)	0% (n=200)
Tg(<i>z-Inta11b-eGFP; cmc2:mCherry</i>)	0% (n=300)
Tg(<i>m-Inta11-eGFP; cmc2:mCherry</i>)	88.9% (n=53)
Mouse Transient Transgenics	
Construct	% of eGFP positive embryos (# eGFP positive / # genotyped positive)
Tg(<i>z-Inta11a-eGFP</i>)	0% (n=0/10)
Tg(<i>z-Inta11b-eGFP</i>)	0% (n=0/7)

Zebrafish stable lines for Tg(*z-Inta11a-eGFP; cmc2:mCherry*); Tg(*z-Inta11b-eGFP; cmc2:mCherry*) were also generated and three genotyped F₁ embryos per line were analysed and confirmed for the absence of *gfp* expression. For Tg(*m-Inta11-eGFP; cmc2:mCherry*), four distinct transgenic lines were also generated and analysed.

Olfactory receptor pseudo-pseudogenes

Lucia L. Prieto-Godino¹, Raphael Rytz^{1,†}, Benoîte Bargeton¹, Liliane Abuin¹, J. Roman Arguello¹, Matteo Dal Peraro^{2,3} & Richard Benton¹

Pseudogenes are generally considered to be non-functional DNA sequences that arise through nonsense or frame-shift mutations of protein-coding genes¹. Although certain pseudogene-derived RNAs have regulatory roles², and some pseudogene fragments are translated³, no clear functions for pseudogene-derived proteins are known. Olfactory receptor families contain many pseudogenes, which reflect low selection pressures on loci no longer relevant to the fitness of a species⁴. Here we report the characterization of a pseudogene in the chemosensory variant ionotropic glutamate receptor repertoire^{5,6} of *Drosophila sechellia*, an insect endemic to the Seychelles that feeds almost exclusively on the ripe fruit of *Morinda citrifolia*⁷. This locus, *D. sechellia* *Ir75a*, bears a premature termination codon (PTC) that appears to be fixed in the population.

However, *D. sechellia* *Ir75a* encodes a functional receptor, owing to efficient translational read-through of the PTC. Read-through is detected only in neurons and is independent of the type of termination codon, but depends on the sequence downstream of the PTC. Furthermore, although the intact *Drosophila melanogaster* *Ir75a* orthologue detects acetic acid—a chemical cue important for locating fermenting food^{8,9} found only at trace levels in *Morinda* fruit¹⁰—*D. sechellia* *Ir75a* has evolved distinct odour-tuning properties through amino-acid changes in its ligand-binding domain. We identify functional PTC-containing loci within different olfactory receptor repertoires and species, suggesting that such ‘pseudo-pseudogenes’ could represent a widespread phenomenon.

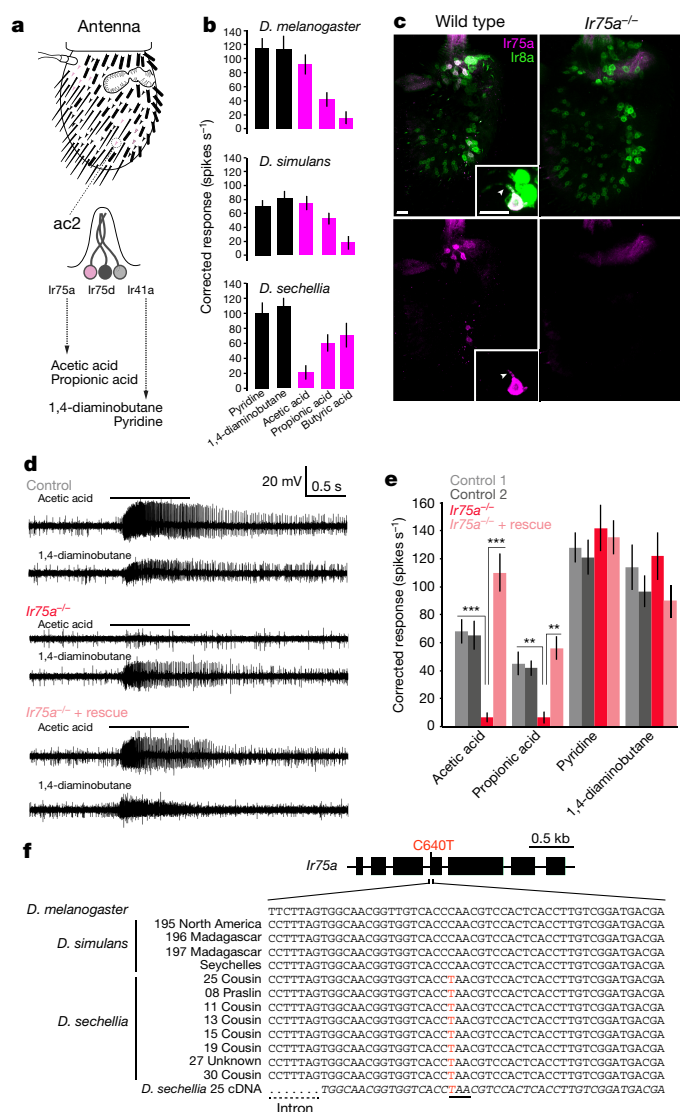


Figure 1 | *Ir75a* encodes an acetic acid receptor in *D. melanogaster* and is a transcribed pseudogene in *D. sechellia*. **a**, Top, schematic of the third antennal segment covered with porous olfactory sensilla of various morphological classes. Bottom, schematic of the ac2 sensillum class, which houses three OSNs that each express different ionotropic receptor genes. **b**, Electrophysiological responses in ac2 sensilla to the indicated odours (mean \pm s.e.m.; mixed genders) in *D. melanogaster*³⁰ ($n = 9$), *D. simulans*³¹ ($n = 9$) and *D. sechellia*³¹ ($n = 8$). The colours of the columns on the histogram distinguishes two broad chemical classes of odours: acids (magenta) and amines (black). **c**, Immunostaining with anti-*Ir75a* (magenta) and anti-*Ir8a* (green) antibodies on antennae of wild-type (left) or *Ir75a*-mutant (*Ir75a*^{MB00253}; right) animals. Insets show the co-localization of *Ir75a* and *Ir8a* in the OSN soma and dendritic compartment (arrowheads). Scale bars, 10 μ m. **d**, Representative traces of extracellular recordings of neuronal responses to the indicated stimuli in ac2 sensilla in control (*Ir75a*-GAL4, *Ir75a*^{MB00253/+}), *Ir75a* hemizygous mutant (*Ir75a*^{MB00253/-}; *Ir75a*^{MB00253}/Df(3L)BSC415) and *Ir75a* rescue (UAS-*D. melanogaster* (Dm)*Ir75a*; *Ir75a*-GAL4, *Ir75a*^{MB00253}/Df(3L)BSC415) animals. Bars above the traces mark 1 s stimulus time. **e**, Quantification of solvent-corrected responses in **d**. Data are mean \pm s.e.m.; mixed genders. Control 1: Df(3L)BSC415/+ ($n = 12$); control 2: *Ir75a*-GAL4, *Ir75a*^{MB00253/+} ($n = 11$); *Ir75a*^{MB00253/-} ($n = 12$); *Ir75a* rescue ($n = 13$). Statistical differences between genotypes were tested using pairwise Wilcoxon rank-sum tests among the responses to each odorant, and p values adjusted for multiple comparisons using the Benjamini–Hochberg method. Significant comparisons to *Ir75a*^{MB00253/-} are shown in the figure (*** $P = 0.0001$, ** $P = 0.001$; for full information about P values see Source Data & Methods). **f**, Top, gene model of *Ir75a* indicating the position of the C640T nucleotide substitution in the *D. sechellia* orthologue. Bottom, genomic sequence spanning this nucleotide position in *D. melanogaster* and several geographically distributed *D. simulans* and *D. sechellia* strains (Methods). The bottom italicized sequence is of the *D. sechellia* cDNA. The *D. sechellia* C640T substitution (red) creates a PTC (underlined).

¹Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland. ²Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland. ³Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland. [†]Present address: Federal Office of Public Health, CH-3003 Bern, Switzerland.

Comparative electrophysiological analysis of olfactory sensory neuron (OSN) responses in closely-related drosophilid species revealed a loss of sensitivity to acetic acid in *D. sechellia* neurons housed in the antennal coeloconic 2 (ac2) sensillum class of sensory hairs (Fig. 1a, b). In *D. melanogaster*, acetic acid is detected by ac2 OSNs expressing *Ir75a*¹¹; these sensilla house two other neurons that are sensitive to amines and express *Ir41a* and *Ir75d*^{11,12} (Fig. 1a). Two lines of evidence support *Ir75a* as the acetic acid receptor. First, the protein is expressed exclusively in these acetic-acid-sensing ac2 neurons, where it co-localizes with the ionotropic receptor co-receptor *Ir8a*¹³ in somata and sensory dendrites (Fig. 1c). Second, protein-null *Ir75a*-mutant animals (Fig. 1c) lack responses to acetic acid (and other organic acids) in these sensilla, while amine ligand-evoked action potentials are unaffected (Fig. 1d, e). Acid sensitivity is restored by expression of an *Ir75a* cDNA in these neurons (Fig. 1d, e).

Ir75a orthologues are present across drosophilids⁶, but *D. sechellia* *Ir75a* is a predicted pseudogene. A C640T nucleotide substitution in the open reading frame (ORF) creates a PTC (CAA→TAA) in exon 4 (Fig. 1f) that is predicted to truncate the protein within the ligand-binding domain (LBD). This PTC is present in all *D. sechellia* strains that we sequenced (from at least two islands of the Seychelles archipelago), but not in any *D. melanogaster* or *D. simulans* strain (Fig. 1f), suggesting that it is a derived change that is fixed in the *D. sechellia* population. We could, however, amplify *Ir75a* cDNA from *D. sechellia* antennal RNA. Sequencing of this cDNA, in addition to data from *D. sechellia* antennal RNA sequencing (Methods), verified that the PTC is not edited or spliced out of the transcript to maintain an intact ORF (Fig. 1f).

The pseudogenization of *D. sechellia* *Ir75a* provided a logical explanation for the loss of responses to acetic acid in this species (Fig. 1b). Nevertheless, *D. sechellia* ac2 sensilla house a neuron that responds to other acidic odours (Fig. 1b), suggesting that another receptor is expressed in these OSNs. We were, however, unable to detect other acid-sensing ionotropic receptors in these cells. We therefore wondered whether the *Ir75a* pseudogene might encode a functional receptor. Indeed, the anti-*Ir75a* antibody stains OSNs in *D. sechellia* with a similar distribution to that of ac2 sensilla (Fig. 2a). Because the epitope of this antibody is encoded upstream of the PTC, we generated a second antibody that recognizes an epitope of the protein that is encoded downstream of the PTC (anti-*Ir75a*^D), and found that it labelled the same cells (Fig. 2a). We also generated a transgene comprising *D. sechellia* *Ir75a* cDNA in which the terminal stop codon was removed and the coding sequence for GFP inserted in-frame with the last coding codon (*DsIr75a:GFP*). As *D. sechellia* is not yet amenable to transgenesis, we expressed this construct in *D. melanogaster* *Ir75a* neurons. GFP fluorescence was detected from *D. sechellia* *Ir75a:GFP* (Fig. 2b and Extended Data Fig. 1), indicating that the PTC is read through, permitting translation of the downstream GFP sequence. No GFP signal was observed with a control construct that retained the terminal stop codon (*DsIr75aSTOP:GFP*) (Fig. 2b and Extended Data Fig. 1).

We next asked whether *D. sechellia* *Ir75a* encodes a functional receptor by misexpressing it in heterologous 'ionotropic receptor decoder' neurons (that is, ac4 *Ir84a*-mutant neurons that lack the endogenous ligand-specific *Ir84a*, but which still express the co-receptor *Ir8a*). In these cells, *D. melanogaster* *Ir75a* endowed sensitivity to acetic and propionic acids (Fig. 2c, d), consistent with the expected endogenous responses of ac2 *Ir75a* OSNs (Fig. 1b). By contrast, *D. sechellia* *Ir75a* conferred responses to propionic, butyric and 2-oxopentanoic acids (Fig. 2c, d). Cluster analysis revealed that the responses of *D. sechellia* *Ir75a* in ionotropic receptor decoder neurons and the endogenous responses of ac2 sensilla neurons group together (Fig. 2e, f). These results provide evidence that the *D. sechellia* *Ir75a* pseudogene encodes a functional olfactory receptor that accounts for the ac2 acid-sensing properties of this species.

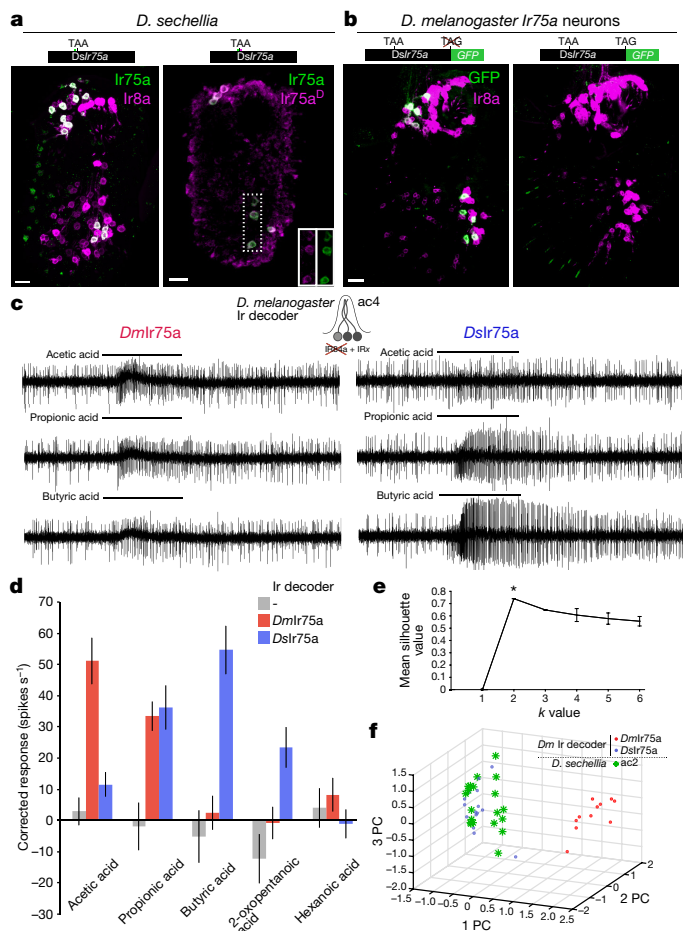


Figure 2 | Translational read-through of the PTC in *D. sechellia* *Ir75a* permits production of a functional olfactory receptor.

a, Immunofluorescence on *D. sechellia* antennae with anti-*Ir75a* (green, recognizing an epitope upstream of the PTC) and anti-*Ir8a* (magenta) (left) or anti-*Ir75a* (green) and anti-*Ir75a*^D (magenta, recognizing an epitope downstream of the PTC) (right). The insets show the separate channels for anti-*Ir75a* and anti-*Ir75a*^D corresponding to the area demarcated with a dashed box. Scale bars, 10 μ m. **b**, Immunofluorescence with anti-GFP antibodies (green) and anti-*Ir8a* antibodies (magenta) on a *D. melanogaster* antenna in which *Ir75a* neurons express transgenes encoding *D. sechellia* (*DsIr75a:GFP* (*UAS-DsIr75a:GFP/+; Ir75a-GAL4/+*) (left) or *Ir75aSTOP:GFP* (*UAS-DsIr75aSTOP:GFP/+; Ir75a-GAL4/+*) (right). **c**, Representative traces of extracellular recordings of neuronal responses to the indicated stimuli in ionotropic receptor (Ir) decoder neurons expressing *D. melanogaster* *Ir75a* (*UAS-DmIr75a; Ir84a^{GAL4}*) or *D. sechellia* *Ir75a* (*UAS-DsIr75a; Ir84a^{GAL4}*). **d**, Quantification of odour-evoked responses in empty ionotropic receptor decoder neurons (*Ir84a^{GAL4}*, $n = 5$), or the decoder neurons expressing *DmIr75a* ($n = 7-11$) or *DsIr75a* ($n = 7-14$) (genotypes as in **c**) (mean \pm s.e.m.; mixed genders). **e**, k -means cluster analysis of the responses of *D. melanogaster* and *D. sechellia* *Ir75a* in the ionotropic receptor decoder and *D. sechellia* ac2 sensilla to the four main agonists used (acetic acid, propionic acid, butyric acid, and 2-oxopentanoic acid). Mean and s.d. of all the solutions within each k value ($n = 100$). The peak silhouette value at $k = 2$ was significantly different from other k values (Student's t -test between $k = 2$ and $k = 3$, $*P < 0.001$), indicating that responses of these three distinct neuron classes statistically fall within two clusters. **f**, Plot of the three first principal components (PC) from a principal component analysis of the same odour response profiles as in **e**.

Reversion of the PTC to the ancestral glutamine-encoding codon (TAA→CAA; *214Q) in transgenic constructs had no effect on expression or function of *D. sechellia* *Ir75a* (Fig. 3a, b and Extended Data Fig. 1), indicating that the PTC is read through efficiently and does not influence odour responses. Translational read-through of

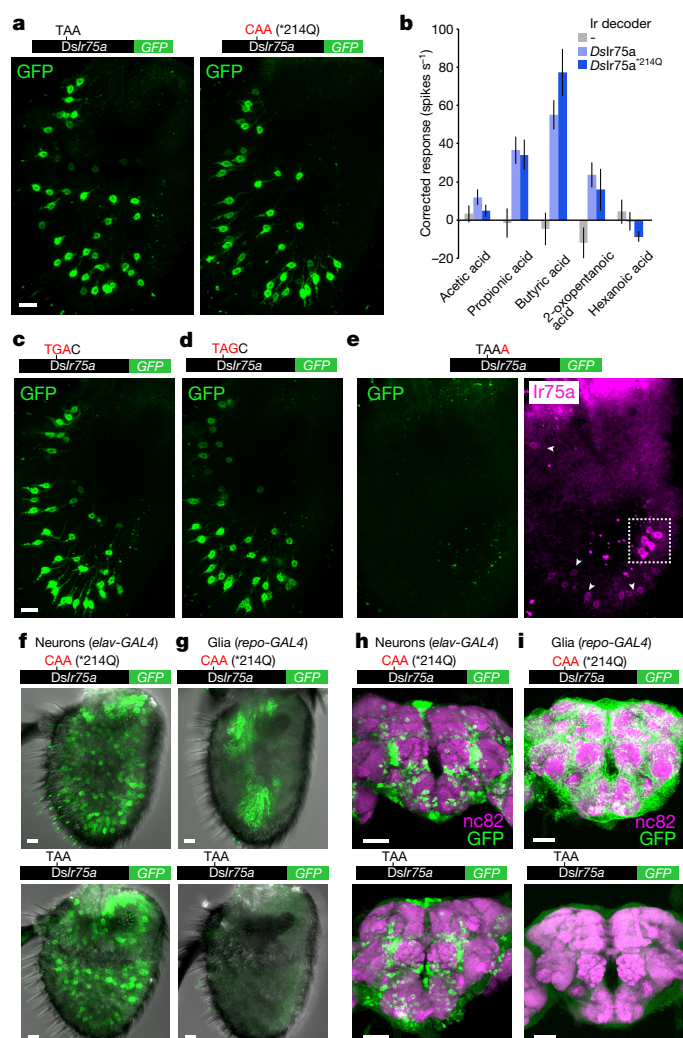


Figure 3 | Efficiency and tissue-specificity of translational read-through of the *D. sechellia* Ir75a PTC. **a**, Immunofluorescence with anti-GFP on *D. melanogaster* antennae in which ionotropic receptor decoder neurons express transgenes encoding *D. sechellia* Ir75a:GFP (UAS-DsIr75a:GFP/+; Ir84a^{GAL4}/+) or Ir75a^{*214Q}:GFP (UAS-DsIr75a^{*214Q}:GFP/+; Ir84a^{GAL4}/+). Scale bar, 10 μ m. **b**, Quantification of odour-evoked responses in empty ionotropic receptor decoder neurons (Ir84a^{GAL4}, $n = 4-5$), or those expressing *D. sechellia* Ir75a (UAS-DsIr75a; Ir84a^{GAL4}, $n = 8-14$) or Ir75a^{*214Q} (UAS-DsIr75a^{*214Q}; Ir84a^{GAL4}, $n = 8-11$). Data are mean \pm s.e.m.; mixed genders. No significant differences were found between the two genotypes in the responses to any of the odours by Student's *t*-test. **c-e**, Immunofluorescence with anti-GFP (green) on *D. melanogaster* antennae in which ionotropic receptor decoder neurons express *D. sechellia* Ir75a:GFP transgenes bearing the indicated mutations to the PTC or 3' nucleotide (genotypes of the form: UAS-DsIr75a^{xxx}:GFP/+; Ir84a^{GAL4}/+). In **e**, immunofluorescence with anti-Ir75a antibodies (magenta) is also shown. Signal is detected in Ir84a neurons with anti-Ir75a (which detects an epitope upstream of the PTC) but not anti-GFP (arrowheads), meaning that this transgene encodes protein up to, but not beyond, the PTC. The dashed white square indicates endogenous Ir75a neurons. Scale bar, 10 μ m. **f**, Immunofluorescence with anti-GFP (green) on *D. melanogaster* antennae in which *elav*-GAL4 drives neuronal expression of *D. sechellia* Ir75a^{*214Q}:GFP (elav-GAL4/+; UAS-DsIr75a^{*214Q}:GFP/+; 8 out of 8 brains were GFP-positive) or Ir75a:GFP (elav-GAL4/+; UAS-DsIr75a:GFP/+; 5 out of 5 GFP-positive). Scale bars, 10 μ m. **g**, Immunofluorescence with anti-GFP (green) on *D. melanogaster* antennae in which *repo*-GAL4 drives glial-specific expression of *D. sechellia* Ir75a^{*214Q}:GFP (UAS-DsIr75a^{*214Q}:GFP/+; repo-GAL4/+; 10 out of 10 were GFP-positive) or Ir75a:GFP (UAS-DsIr75a:GFP/+; repo-GAL4/+; 0 out of 10 were GFP-positive). Scale bars, 10 μ m. **h**, Immunofluorescence with anti-GFP (green) and synaptic neuropil marker nc82 (magenta) antibodies on *D. melanogaster* brains in which *elav*-GAL4 drives neuronal expression of *D. sechellia* Ir75a^{*214Q}:GFP (6 out of 6 were GFP-positive) or Ir75a:GFP (8 out of 8 were GFP-positive); genotypes as in **f**. Scale bars, 50 μ m. **i**, Immunofluorescence with anti-GFP (green) and anti-nc82 (magenta) on *D. melanogaster* brains in which *repo*-GAL4 drives glial-specific expression of *D. sechellia* Ir75a^{*214Q}:GFP (6 out of 6 were GFP-positive) or Ir75a:GFP (0 out of 8 were GFP-positive); genotypes as in **g**. Scale bars, 50 μ m.

terminal stop codons, resulting in C-terminal extensions, has been characterized for several eukaryotic genes^{15,16}; in these cases, the 'leakiness' of translation arrest is predicted to depend on the termination codon (TGA>TAA>TAG) and the immediate 3' nucleotide (C>T>G>A)^{14,15}. We investigated the *cis*-regulatory elements that determine the high efficiency of read-through of the Ir75a PTC—which has the second most leaky termination codon context (TAAC)¹⁴—by generating additional read-through GFP reporters bearing mutations in this sequence (Fig. 3c, e). Replacement of the TAA PTC with either TGA or TAG did not affect GFP expression (Fig. 3c, d and Extended Data Fig. 1). By contrast, replacing the immediate 3' cytosine nucleotide with an adenosine almost completely blocked GFP expression (Fig. 3e and Extended Data Fig. 1), although this transgene still produces a truncated protein, as detected by the anti-Ir75a antibody (Fig. 3e). These results indicate that sequence context outside, but not within, the Ir75a PTC is critical for determining read-through efficiency.

The expression of full-length protein by PTC-containing *D. sechellia* Ir75a transgenes in two populations of *D. melanogaster* OSNs (that is, Ir75a and ionotropic receptor decoder neurons) indicates that the mechanisms that permit read-through are not species- or OSN-class-specific. We investigated whether read-through occurs in other cell types by using an *actin5C*-GAL4 driver to broadly express *D. sechellia* Ir75a:GFP and, as a control, *D. sechellia* Ir75a^{*214Q}:GFP. *D. sechellia* Ir75a^{*214Q}:GFP was detected in many, but not all, cells (Extended Data Fig. 2); this heterogeneity may arise from the variable expression of *actin5C*-GAL4 in different cell types.

Nevertheless, the GFP-positive cells encompassed neurons and non-neuronal support cells (Extended Data Fig. 2, arrowheads). By contrast, *D. sechellia* Ir75a:GFP was detected exclusively in neurons (Extended Data Fig. 2). To confirm this finding, we compared the expression of *D. sechellia* Ir75a^{*214Q}:GFP and Ir75a:GFP transgenes induced either by a pan-neuronal (*elav*-GAL4) or pan-glial (*repo*-GAL4) driver. Both transgenes produced similar GFP signals in sensory neurons throughout the antenna (Fig. 3f). However, only Ir75a^{*214Q}:GFP was detectable in glia (Fig. 3g). Similarly, we detected broad neuronal expression of both Ir75a^{*214Q}:GFP and Ir75a:GFP in the brains of these animals (Fig. 3h), but only Ir75a^{*214Q}:GFP was expressed in glia (Fig. 3i). Thus, efficient read-through of the *D. sechellia* Ir75a PTC occurs in diverse neuronal classes, but not in non-neuronal cells.

We next investigated the molecular basis of the different ligand-response profiles of *D. sechellia* and *D. melanogaster* Ir75a. As Ir75a-dependent neuron responses are conserved between *D. melanogaster* and *D. simulans* (Fig. 1b), we reasoned that ligand-specificity determinants would be conserved in Ir75a sequences from these species but differ in *D. sechellia* Ir75a; of eight such positions (excluding the PTC), six lie within the predicted bi-lobed LBD (Extended Data Fig. 3). Three of these positions are located within the putative ligand-binding pocket in a protein homology model of the *D. sechellia* Ir75a LBD (Fig. 4a). We tested their function by generating a version of *D. melanogaster* Ir75a in which each of these positions was mutated to encode the amino acid present in *D. sechellia* Ir75a, and expressed this mutant (*D. melanogaster* Ir75a^{T289S,Q536K,F538L}).

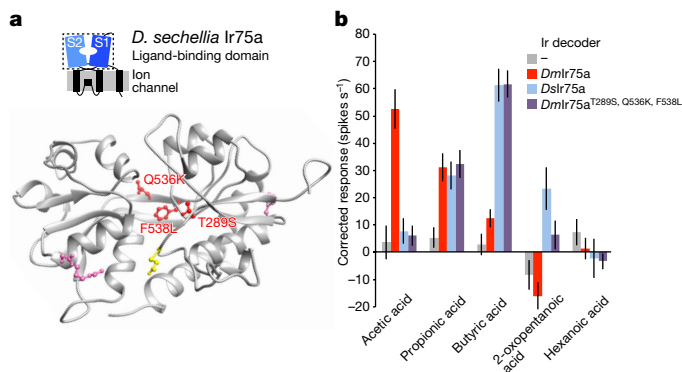


Figure 4 | Molecular basis of the functional divergence of *D. sechellia* Ir75a. **a**, Protein homology model of the LBD of *D. sechellia* Ir75a (Methods). The side chains of the residues that are different in *D. sechellia* Ir75a compared to *D. simulans* and *D. melanogaster* Ir75a are represented in pink, and the subset of these mutated in this study are shown in red. The location of the residue putatively encoded by the PTC is shown in yellow. **b**, Quantification of odour-evoked responses in empty ionotropic receptor decoder neurons (*Ir84a^{GAL4}*, $n = 5-6$) or the decoder neurons expressing *D. melanogaster* Ir75a ($n = 9$), *D. sechellia* Ir75a ($n = 8-9$) or the *D. melanogaster* Ir75a^{T289S,Q536K,F538L} mutant ($n = 14$) (genotypes used, *UAS-DxIr75a^{xxx};Ir84a^{GAL4}*) (mean \pm s.e.m, mixed genders); experiments for all transgenes were performed in parallel. Responses to each odour of *D. sechellia* Ir75a and *D. melanogaster* Ir75a^{T289S,Q536K,F538L} are statistically indistinguishable (Wilcoxon rank-sum test).

in ionotropic receptor decoder neurons. This engineered receptor conferred responses indistinguishable from those of *D. sechellia* Ir75a (Fig. 4b), indicating the importance of one or more of these residues as odour-specificity determinants.

Finally, we sought to identify other functional olfactory receptor pseudogenes. Within wild-caught isolates of *D. melanogaster* (Methods), we identified several strains in which the *Ir75b* ORF contains a C517T substitution, creating a PTC (Fig. 5a and Methods). Immunostaining of antennae from flies of these lines with an Ir75b antibody (recognizing an epitope encoded downstream of the PTC) showed a pattern comparable to controls (Fig. 5a, b), indicating that this PTC is read through. Consistently, electrophysiological recordings of Ir75b neurons in one of these lines (Raleigh707, RAL707) revealed robust responses when presented with known agonists¹¹ (Fig. 5c). In another strain (RAL441), the *Ir31a* ORF contains a T1805G substitution that is predicted to truncate the receptor before the third transmembrane domain (Fig. 5d). Nevertheless, electrophysiological recordings of RAL441 Ir31a neurons stimulated with 2-oxopentanoic acid¹¹, revealed clear responses (Fig. 5e). We also identified a segregating PTC-containing allele of the odorant receptor gene *Or35a* (in Tasmanian strains T09 and T29) (Fig. 5f). In strain T09, responses of Or35a neurons to octanol¹² were readily detected (Fig. 5g). These findings indicate that the phenomenon of functional olfactory pseudogenes is restricted neither to a particular species nor to a specific receptor repertoire.

Our efforts to understand the molecular basis of the loss of olfactory sensitivity to acetic acid in *D. sechellia* led us to discover a notable and, to our knowledge, unprecedented evolutionary trajectory of a presumed pseudogene. Efficient read-through of a PTC in *D. sechellia* Ir75a permits production of a full-length receptor protein, in which reduction in acetic acid sensitivity and gain of responses to other acids is due to lineage-specific amino acid substitutions in the LBD pocket. The PTC does not noticeably influence the activity of *D. sechellia* Ir75a, suggesting that it is selectively neutral from an evolutionary standpoint. We propose that it became fixed through genetic drift, given *D. sechellia*'s persistent low effective population size¹⁷.

It is not yet clear how the *D. sechellia* Ir75a PTC is read through. It cannot be because of insertion of the alternative amino acid

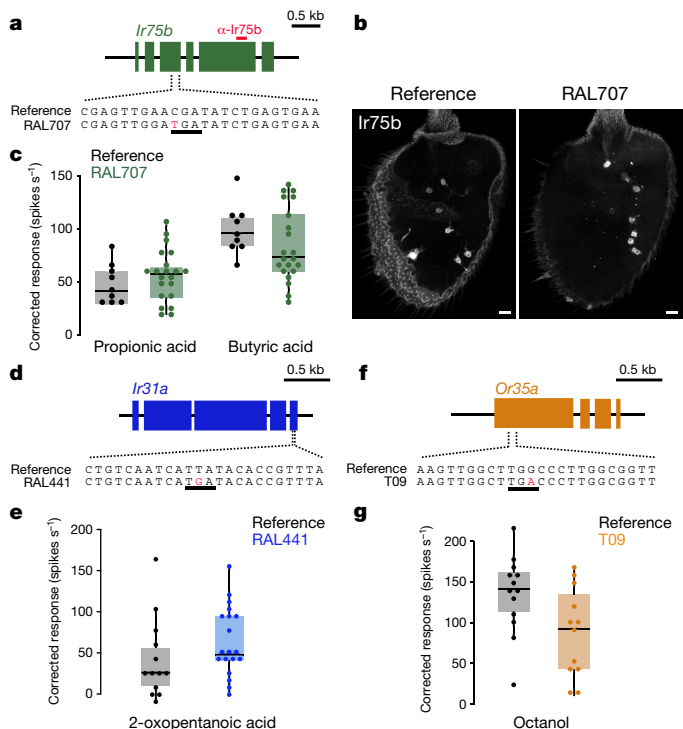


Figure 5 | Functional pseudogene alleles of other receptors in other species. **a**, Gene structure of *D. melanogaster* Ir75b indicating the position of the PTC in the RAL707 strain (an identical sequence is found in all other strains containing a PTC at this position; see Methods). The region encoding the epitope recognized by anti-Ir75b antibodies is indicated with a red bar. **b**, Immunofluorescence with anti-Ir75b antibodies on antennae of reference *D. melanogaster* (Methods) and RAL707 flies. Scale bars, 10 μ m. **c**, Quantification of odour-evoked responses in ac3 sensilla to propionic acid and butyric acid in control (reference *D. melanogaster*) and RAL707 flies. Box plots indicate the median and first and third quartile of the data; each dot corresponds to one recording. **d**, Gene structure of *D. melanogaster* Ir31a, indicating the position of the PTC in the RAL441 strain. **e**, Quantification of odour-evoked responses to 2-oxopentanoic acid in ac1 sensilla of control (reference *D. melanogaster*) and RAL441 flies. **f**, Gene structure of *D. melanogaster* Or35a, indicating the position of the PTC in the T09 (and T29) strain. **g**, Quantification of odour-evoked responses to octanol in ac3 sensilla of control (reference *D. melanogaster*) and T09 flies.

selenocysteine (which is incorporated at UGA¹⁸). Moreover, no suppressor tRNAs are known in *D. melanogaster*¹⁹ and ribosomal frame-shifting is also unlikely because there is no change in the reading frame after the PTC. We suggest that read-through is due to PTC recognition by a near-cognate tRNA that allows insertion of an amino acid instead of translation termination. Although the *trans*-acting factors regulating read-through are unclear, the neuronal specificity of this process is reminiscent of RNA editing and micro-exon splicing, in which key responsible regulatory proteins are neuronally enriched^{20,21}. We therefore speculate that tissue-specific expression differences in tRNA populations underlie neuron-specific read-through.

Although it has long been known that viruses display PTC read-through²², case studies in eukaryotes are largely limited to artificial scenarios in which nonsense mutations have been introduced through random or site-directed mutagenesis²³⁻²⁵, or in human disease-causing alleles with low read-through rates²⁶. Our characterization of four protein-coding PTC-containing genes demonstrates that the read-through of naturally occurring PTCs can be sufficiently efficient to permit the functionality of pseudogenes and the maintenance of these variants in populations. This finding further highlights the plasticity of translational regulation, allowing for the phenotypic buffering of genetic changes²⁷. It should also prompt the experimental examination

of the hundreds of PTC-containing presumed pseudogenes, both within and beyond chemosensory gene families in insects²⁸, humans²⁹ and other organisms.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 6 September 2016.

Published online 24 October 2016.

- Salmena, L. Pseudogene redux with new biological significance. *Methods Mol. Biol.* **1167**, 3–13 (2014).
- Poliseno, L. *et al.* A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature* **465**, 1033–1038 (2010).
- Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5'UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).
- Nei, M., Niimura, Y. & Nozawa, M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. *Nat. Rev. Genet.* **9**, 951–963 (2008).
- Benton, R., Vannice, K. S., Gomez-Diaz, C. & Vosshall, L. B. Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149–162 (2009).
- Croset, V. *et al.* Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, e1001064 (2010).
- Stensmyr, M. C. *Drosophila sechellia* as a model in chemosensory neuroecology. *Ann. NY Acad. Sci.* **1170**, 468–475 (2009).
- Gorter, J. A. *et al.* The nutritional and hedonic value of food modulate sexual receptivity in *Drosophila melanogaster* females. *Sci. Rep.* **6**, 19441 (2016).
- Becher, P. G., Bengtsson, M., Hansson, B. S. & Witzgall, P. Flying the fly: long-range flight behavior of *Drosophila melanogaster* to attractive odors. *J. Chem. Ecol.* **36**, 599–607 (2010).
- Farine, J. P., Legal, L., Moreteau, B. & Le Quere, J. L. Volatile components of ripe fruits of *Morinda citrifolia* and their effects on *Drosophila*. *Phytochemistry* **41**, 433–438 (1996).
- Silbering, A. F. *et al.* Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *J. Neurosci.* **31**, 13357–13375 (2011).
- Yao, C. A., Ignell, R. & Carlson, J. R. Chemosensory coding by neurons in the coelomic sensilla of the *Drosophila* antenna. *J. Neurosci.* **25**, 8359–8367 (2005).
- Abuin, L. *et al.* Functional architecture of olfactory ionotropic glutamate receptors. *Neuron* **69**, 44–60 (2011).
- Jungreis, I. *et al.* Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.* **21**, 2096–2113 (2011).
- Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R. & Weissman, J. S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *eLife* **2**, e01179 (2013).
- Namy, O. *et al.* Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **31**, 2289–2296 (2003).
- Legrand, D., Vautrin, D., Lachaise, D. & Cariou, M. L. Microsatellite variation suggests a recent fine-scale population structure of *Drosophila sechellia*, a species endemic of the Seychelles archipelago. *Genetica* **139**, 909–919 (2011).
- Shchedrina, V. A. *et al.* Analyses of fruit flies that do not express selenoproteins or express the mouse selenoprotein, methionine sulfoxide reductase B1, reveal a role of selenoproteins in stress resistance. *J. Biol. Chem.* **286**, 29449–29461 (2011).
- Chan, P. P. & Lowe, T. M. tRNADB: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
- Palladino, M. J., Keegan, L. P., O'Connell, M. A. & Reenan, R. A. dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* **6**, 1004–1018 (2000).
- Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
- Namy, O. & Rousset, J. P. in *Recoding: Expansion of Decoding Rules Enriches Gene Expression* (eds Atkins, J. F. & Gesteland, R. F.) 79–100 (Springer, 2010).
- Kopczynski, J. B., Raff, A. C. & Bonner, J. J. Translational readthrough at nonsense mutations in the *HSF1* gene of *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **234**, 369–378 (1992).
- Washburn, T. & O'Tousa, J. E. Nonsense suppression of the major rhodopsin gene of *Drosophila*. *Genetics* **130**, 585–595 (1992).
- Samson, M. L., Lisbin, M. J. & White, K. Two distinct temperature-sensitive alleles at the *elav* locus of *Drosophila* are suppressed nonsense mutations of the same tryptophan codon. *Genetics* **141**, 1101–1111 (1995).
- Keeling, K. M., Xue, X., Gunn, G. & Bedwell, D. M. Therapeutics based on stop codon readthrough. *Annu. Rev. Genomics Hum. Genet.* **15**, 371–394 (2014).
- Jagannathan, S. & Bradley, R. K. Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res.* <http://dx.doi.org/10.1101/gr.205070.116> (2016).
- Huang, W. *et al.* Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.* **24**, 1193–1208 (2014).
- Pei, B. *et al.* The GENCODE pseudogene resource. *Genome Biol.* **13**, R51 (2012).
- Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).

Acknowledgements We acknowledge C. Carracedo, P. Casares, the Bloomington *Drosophila* Stock Center (NIH P40OD018537), the *Drosophila* Species Stock Center (UCSD), and the Developmental Studies Hybridoma Bank (NICHHD of the NIH, University of Iowa) for reagents. We thank members of the Benton laboratory for discussions and comments on the manuscript. L.L.P.-G. was supported by a FEBS long-term fellowship; R.R. was supported by a Roche Research Foundation fellowship. J.R.A. was supported by a post-doctoral fellowship from Novartis Foundation for medical-biological research (12A14). M.D.P.'s laboratory was supported by the SNSF. Research in R.B.'s laboratory was supported by ERC Starting Independent Researcher and Consolidator Grants (205202 and 615094), an HFSP Young Investigator Award (RGY0073/2011) and the SNSF Nano-Tera Envirobot project (20NA21_143082).

Author Contributions L.L.P.-G. and R.B. conceived the project. L.L.P.-G., R.R., B.B. and M.D.P. designed experiments. L.L.P.-G., R.R., L.A. and R.B. generated reagents. L.L.P.-G. performed electrophysiology, bioinformatics and histology. R.R. performed electrophysiology and molecular analyses. B.B. generated protein models. L.A. performed histology. J.R.A. performed bioinformatics. L.L.P.-G., R.R., B.B., J.R.A. and R.B. analysed data. L.L.P.-G. and R.B. wrote the paper with input from all authors.

Author Information Sequences of cloned ORFs have been deposited in GenBank under accession numbers: KX694388 (*D. melanogaster* *Ir75a*) and KX694389 (*D. sechellia* *Ir75a*). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. (richard.benton@unil.ch).

Reviewer Information *Nature* thanks A. Jacobson, M. Stensmyr and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. These experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Molecular biology and transgenesis. *D. melanogaster* and *D. sechellia* Ir75a ORFs (including the PTC in *D. sechellia* Ir75a) were cloned into pUAST-attB³². Site-directed mutagenesis of *D. sechellia* Ir75a and *D. melanogaster* Ir75a, reverse-transcription PCR amplification and sequencing of genomic amplicons to verify the presence of the *D. sechellia* Ir75a PTC were performed using standard procedures. For the transgene in which the nucleotide 3' of the PTC was mutated (Fig. 3e), we changed codon 215 from CGT→AGA to maintain the identity of the encoded amino acid (arginine). Oligonucleotide and plasmid sequences are available upon request. New transgenes were integrated in attP40 using the phiC31 site-specific integration system³² by Genetic Services Inc. or BestGene Inc. All transgenes were sequence-verified both before and after integration into *D. melanogaster*.

Drosophila strains. Flies were maintained at 25 °C in 12 h light:12 h dark conditions. *D. melanogaster* wild-type refers to w¹¹¹⁸, unless noted otherwise. *D. sechellia* wild-type was 14021-0248.25 (*Drosophila* Species Stock Center, UCSD). The sequenced region of Ir75a shown in Fig. 1f was amplified from the following strains (from the *Drosophila* Species Stock Center, unless noted otherwise): *D. sechellia*: 14021-0248.08, 14021-0248.11, 14021-0248.13, 14021-0248.15, 14021-0248.19, 14021-0248.25, 14021-0248.27, 14021-0248.30; *D. simulans*: 14021-0251.195, 14021-0251.196, 14021-0251.197, as well as a Seychelles-isolated *D. simulans*³³. Other published mutant and transgenic lines used were: *Ir84a*^{GAL4} (ref. 34), *Ir75a-GAL4* (ref. 11), *UAS-CD8:GFP³⁵*, *Mi[ET1]Ir75a[MB00253]³⁶*, *Df(3L)BSC415* (*Ir75a* deficiency)³⁷, *actin5C-GAL4* (ref. 38), *elav-GAL4* (Bloomington 458), *repo-GAL4* (Bloomington 7415), *RAL441*, *RAL707* (ref. 39) and *Tasmania T09* (ref. 40).

Sequence analysis. We downloaded *D. sechellia* antennal RNA-sequencing datasets⁴¹ from the NCBI Gene Expression Omnibus repository (GEO accessions GSE67861 and GSE67587; SSR files SRR1952772, SRR1952777, SRR1973487, SRR1973490). The sra files were converted to fastq files and remapped to the *D. sechellia* genome (r1.3) using TopHat (v2.0.13; -b2-sensitive). The genomic index and splicing index were also generated with TopHat using the *D. sechellia* gtf (r1.3). The resulting bam files were visualized within IGV (v2.3.63), and we manually inspected reads that covered the Ir75a PTC. Within these four datasets, ~100% of the reads supported the presence of the PTC-causing 'T' allele (only 6/1777 reads within all four datasets supported an alternative nucleotide, within the noise of sequencing errors).

PTCs in other olfactory receptor genes were identified in the *Drosophila melanogaster* Genetic Reference Panel (DGRP; <http://dgrp2.gnets.ncsu.edu/>)³⁹ and/or the Global Diversity Lines (GDL)^{40,42}. For Ir75b, the following lines contain the same PTC: DGRP (RAL181, RAL189, RAL227, RAL320, RAL348, RAL352, RAL358, RAL374, RAL379, RAL382, RAL385, RAL395, RAL399, RAL439, RAL461, RAL531, RAL596, RAL707, RAL712, RAL716, RAL730, RAL804, RAL805, RAL821, RAL855, RAL884), GDL (B10, B11, B12, I26, N01, N02, N07, N25, T29). We re-sequenced the PTC-containing region and confirmed Ir75b protein expression in all strains from the GDL (data not shown). For Ir31a, only RAL441 contains the PTC. For Or35a, GDL Tasmanian strains T09 and T29 contain the same PTC.

Histology and morphological analyses. Immunofluorescence on whole-mount antennae or antennal cryosections (used only in Fig. 1c, inset, and Fig. 2a), were performed as described^{11,13,43}. Affinity-purified antibodies were generated by Proteintech Group, Inc., against the following peptides: KRKYGNREQLTDVVLVRV (anti-Ir75a, in rabbits; used at 1:100 for whole-mount antennae, and 1:500 for cryosections), and RPLTSLDDELIRLSQEND (anti-Ir75a^D, in guinea pigs; used at 1:10) and PDVRDLYRKKVLGSKRSPD (anti-Ir75b, in guinea pigs; used at 1:500); these peptides are predicted antigenic, surface-exposed sequences conserved in *D. melanogaster*, *D. simulans* and *D. sechellia* orthologues. Other antibodies used were (concentrations listed are for whole-mount antennal stains, for cryosections, antibodies were used at tenfold higher dilutions): guinea-pig anti-Ir8a 1:1,000 (ref. 13), chicken anti-GFP 1:100 (AbCam), rat anti-Elav 1:10 (7E8A10; Developmental Studies Hybridoma Bank) and mouse monoclonal nc82 1:10 (Developmental Studies Hybridoma Bank). Alexa488-, Cy3- and Cy5-conjugated goat anti-mouse IgG, goat anti-guinea-pig IgG and goat anti-rabbit IgG secondary antibodies (Molecular Probes and Jackson Immunoresearch) were used at 1:100, and Alexa488-conjugated donkey anti-chicken IgG secondary antibodies were used at 1:500.

The quantifications in Extended Data Fig. 1 were performed using ImageJ (<http://imagej.nih.gov/ij/>). In brief, for each antenna a single plane with several

Ir75a-positive cells was chosen and cropped to exclude background from the surroundings. The Ir75a channel was used to create a mask of the cells by using the auto-threshold function. This mask was then applied to the composite image (with overlapping anti-Ir75a in the red channel and anti-GFP in the green channel). The new image was analysed using the 'color histogram' tool to give the total number of red and green pixels within the masked cells; these values were used to calculate the ratio of green to red pixels.

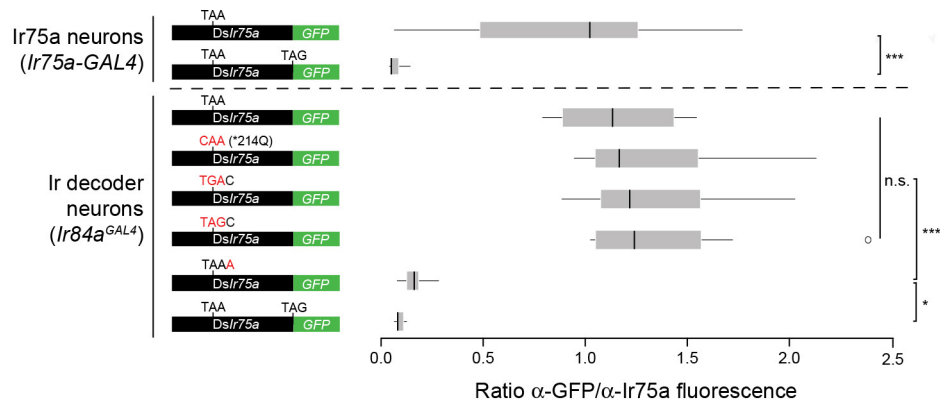
Electrophysiology. Single sensillum electrophysiological recordings were performed essentially as described^{44,45} on 2–10 day old animals. The sample sizes (*n*) indicated in the figure legends correspond to biological replicates (different sensilla), with a maximum of three sensilla per animal. Exact sample sizes for each experimental or group condition are provided in the Source Data. Genotypes (not blinded to experimenter) were interleaved to minimise effects of time-of-day and animal age. All odour-evoked responses were corrected for solvent-evoked spikes. Chemicals were purchased from Sigma-Aldrich and were of the highest purity available. Odorants were used at 1% (v/v) in all experiments unless otherwise noted in the figure legends. Odour stimulus cartridges (10 µl odour dilution on ~5 × 5 mm Sugi strip placed in a 2-ml plastic syringe) were prepared freshly before each recording session; cartridges were interleaved, with a maximum of five uses. Stimuli, with CAS number and solvents used in brackets, are as follows: 1,4-diaminobutane (110-60-1; H₂O), 2-oxopentanoic acid (1821-02-9; paraffin oil), acetic acid (64-19-7; H₂O), butyric acid (107-92-6; H₂O), hexanoic acid (142-62-1; H₂O), octanol (111-87-5; paraffin oil), phenylethylamine (64-04-0; paraffin oil), propionic acid (79-09-4; H₂O), pyridine (110-86-1; paraffin oil). Ionotropic receptor decoder neurons are ac4 *Ir84a*-mutant neurons that lack the endogenous ligand-specific Ir84a, but that still express the co-receptor Ir8a³⁴. For measurement of odour-evoked responses of Ir75b neurons in ac3 sensilla (Fig. 5c), we note approximately half of the analysed sensilla belong to 'ac3II' class (expressing Ir75c) that are electrophysiologically indistinguishable from ac3I sensilla (expressing Ir75b; L.L.P.-G., R.R. and R.B., unpublished data). Thus, if the RAL707 PTC-bearing allele is non-functional, we would expect half of the sensilla to show no responses, which is not the case.

Statistical analysis. Sample sizes were fixed before data analysis, based on preliminary studies. Data were analysed and plotted using 'R project' (<http://R-project.org>). Data were analysed statistically using the Shapiro–Wilk test to assess for normality followed by a two-tailed Student's *t*-test or a Wilcoxon rank-sum test as appropriate. When a *P* value correction for multiple comparisons was needed, the Benjamini–Hochberg method was used. Full statistical test results are provided in the Source Data files for each figure.

Cluster analysis of odour responses. Spike-count data (from responses of endogenous or ionotropic receptor decoder neurons) were imported to Matlab (Mathworks), and standardized using z-scores across each recording. An unbiased *k*-means cluster analysis was performed using the response properties to the indicated number of odours. The optimal number of clusters was determined by the silhouette method⁴⁶. The silhouette value is a measure of both how tightly a data point is associated with its assigned cluster and how dissimilar it is from other clusters, and it peaks at the 'correct' number of clusters. If the distribution is unimodal (that is, all data fall within one cluster) the silhouette value does not peak and is similar for each value of *k*. In brief, we first ran iteratively the Matlab *k*-means algorithm 100 times for *k* values between 2 and 6. We then calculated the silhouette value for each *k*-means solution, and subsequently computed and plotted the mean silhouette value and its standard deviation of all of the solutions within each *k* value. The results of the clusters from the *k*-mean analysis were plotted in Matlab on the principal component space after performing a principal component analysis using the same odour dataset. The scripts used to analyse data are available upon request.

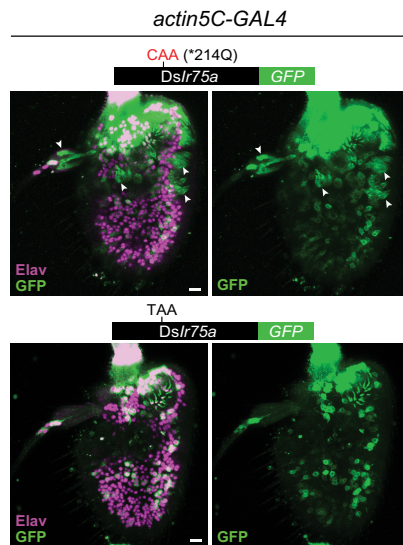
Protein homology modelling. A multiple-sequence alignment of the LBD of *D. sechellia* Ir75a, *D. melanogaster* antennal ionotropic receptors (Ir8a, Ir21a, Ir25a, Ir31a, Ir40a, Ir41a, Ir64a, Ir75a, Ir75b, Ir75c, Ir76a, Ir76b, Ir84a, Ir92a, Ir93a)⁶, *Rattus norvegicus* GluK2 (UniProt ID P42260), *Rattus norvegicus* GluA2 (P19491), *Adineta vaga* GluR1 (E9P5T5) and *Synechocystis* PCC 6803 GluR0 (P73797) was generated by PROMALS3D⁴⁷. A GT dipeptide sequence was introduced between the S1 and S2 domains of the *Drosophila* proteins to facilitate alignment with the linker sequence included in the crystallized GluA2, GluK2, AvGluR1 and GluR0 LBDs. Alignment of Ir75a proteins was curated according to PSIPRED secondary structure predictions⁴⁸. Models of the *D. sechellia* Ir75a LBD (V205–T318–GT–K434–C579) were built using MODELLER (mod9.12)⁴⁹ using as templates the apo and ligand-bound crystal structures of GluA2 (PDB ID: 1FTO apo state; 1FTM ligand bound⁵⁰). The results of standard MODELLER energy functions, molpdf and DOPE, were highly similar between generated models. We illustrate in Fig. 4a the model with the lowest DOPE energy function score.

32. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
33. Carracedo, M. C., Asenjo, A. & Casares, P. Genetics of *Drosophila simulans* male mating discrimination in crosses with *D. melanogaster*. *Heredity* **91**, 202–207 (2003).
34. Grosjean, Y. *et al.* An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature* **478**, 236–240 (2011).
35. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
36. Bellen, H. J. *et al.* The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167**, 761–781 (2004).
37. Cook, R. K. *et al.* The generation of chromosomal deletions to provide extensive coverage and subdivision of the *Drosophila melanogaster* genome. *Genome Biol.* **13**, R21 (2012).
38. Ito, K., Awano, W., Suzuki, K., Hiromi, Y. & Yamamoto, D. The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124**, 761–771 (1997).
39. Mackay, T. F. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**, 173–178 (2012).
40. Grenier, J. K. *et al.* Global diversity lines—a five-continent reference panel of sequenced *Drosophila melanogaster* strains. *G3 (Bethesda)* **5**, 593–603 (2015).
41. Shiao, M. S. *et al.* Expression divergence of chemosensory genes between *Drosophila sechellia* and its sibling species and its implications for host shift. *Genome Biol. Evol.* **7**, 2843–2858 (2015).
42. Arguello, J. R. *et al.* Extensive local adaptation within the chemosensory system following *Drosophila melanogaster*'s global expansion. *Nat. Commun.* **7**, ncomms11855 (2016).
43. Saina, M. & Benton, R. Visualizing olfactory receptor expression and localization in *Drosophila*. *Methods Mol. Biol.* **1003**, 211–228 (2013).
44. Benton, R., Vannice, K. S. & Vosshall, L. B. An essential role for a CD36-related receptor in pheromone detection in *Drosophila*. *Nature* **450**, 289–293 (2007).
45. Benton, R. & Dahanukar, A. Electrophysiological recording from *Drosophila* olfactory sensilla. *Cold Spring Harb. Protoc.* **2011**, 824–838 (2011).
46. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: an Introduction to Cluster Analysis*. (Wiley-Interscience, 2005).
47. Pei, J., Kim, B. H. & Grishin, N. V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res.* **36**, 2295–2300 (2008).
48. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
49. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
50. Armstrong, N. & Gouaux, E. Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. *Neuron* **28**, 165–181 (2000).



Extended Data Figure 1 | Quantification of efficiency and tissue-specificity of translational read-through of the *D. sechellia* *Ir75a* PTC. Quantification of GFP staining in the cell bodies of neurons expressing different read-through reporter constructs in different populations of OSNs (see Figs 2, 3 for genotypes). GFP fluorescence levels were

normalized by anti-Ir75a fluorescence levels in the Cy3 channel within each analysed cell. Box plots indicate the median and first and third quartile of the data. * $P < 0.05$, *** $P < 0.0005$, not significant (n.s.) $P > 0.05$ (all data analysed using pairwise Wilcoxon rank-sum test, Benjamini–Hochberg correction).



Extended Data Figure 2 | Tissue specificity of translational read-through of the *D. sechellia* *Ir75a* PTC. Immunofluorescence with anti-GFP (green) and the neuron nuclear marker anti-Elav (magenta) on whole-mount *D. melanogaster* antennae in which *actin5C-GAL4* drives broad expression of *D. sechellia* *Ir75a*^{*214Q}:GFP (*UAS-DsIr75a*^{*214Q}:GFP/*act5C-GAL4*) or *Ir75a*:GFP (*UAS-DsIr75a*:GFP/*act5C-GAL4*). Arrowheads indicate examples of GFP-expressing, Elav-negative, non-neuronal cells that were observed in 6 out of 6 antennae expressing the control transgene lacking the PTC, and in 0 out of 6 antennae expressing the PTC-containing transgene. Note that the neuronal GFP signal of both transgenes is heterogeneous across the antenna, possibly because of the variable strength of driver expression and/or instability of the GFP-tagged receptors in heterologous neurons. Scale bars, 10 μ m.

Dmel MQLVQLANFVLDNLVQSRIGFIVLFHCWQSDSELKFAQQFMKPIHPILVYHQFVQMRGVLNWSHLELSYMGHTQPTLAI 79
Dsim MQLVQLANFVLDNLVQSRIGFIVLFHCWQSDSELKFAEQFMKPIHPILVYHQFVQMRGVHNSHLELNMGHTQPTLAI
Dsec MQLVQLANFVLDNLVQSRIGFIVLFHCWQSDSELKFAEQFMKPIHPILVYHQFVQMRGVLNWSHLELNMGHTQPTLAI

80 YVDIKCDQTQDLLEASREQIYNQHYHWLLVGNQSKLEFYDLFGLFNISIDADVSYVKEIQDNNDSVAYAVHDVYNNG 158
 YVDRKCDQAQDLLEASREQIYNQHYHWLLVGNQSELEFNDLFALFNISIDADVSYVKEIQDNNDSVAYAVDYNNNG
 YVDIKCDQAQDLLEASREQIYNQHYHWLLVGNQSELEFNDLFALFNISIDADVSYVKEIQDNNDSVAYAVDYNNNG
 α-Ir75a: S₁ α-Ir75a^h

159 KIIGGQLNVTGSHEMSCDPFVCRTRHLSLQKRSKYGNREQLTDVVLRVATVVVTRPRLTSSDDELIRFLSQENDTHID 237
 KIIGGQLNVTGSHEMSCDPFECRTRHLSLQKRSKYGNREQLTDVVLRVATVVVTRPRLTSSDDELIRFLSQENDTHID
 KIIGGQLNVTGSHEMSCDPFKCRTRHLSLQKRSKYGNREQLTDVVLRVATVVVTRPRLTSSDDELIRFLSQENDTHID

238 SLARFGFHLTLILRDLLHCKMKFIFSDSWSKSDVVGSGVAVVDQTADLTATPSLATEGRLKYLSAIETGFFRSVCIF 316
 SLARFGFHLTLILRDLLHCKMKFIFSDSWSKSDVVGSGVAVVDQTADLTATPSLATEGRLKYLSAIETGFFRSVCIF
 SLARFGFHLTLILRDLLHCKMKFIFSDSWSKSDVVGSGVAVVDQTADLTATPSLATEGRLKYLSAIETGFFRSVCIF

317 RTPHNAGLRGDVFLQPFSPVWYLFGGVLSLIGVLLWITFYMECKRMQKRWRDLPLSLLSTFLISFGAACIQSSSLIP 395
 RTPHNAGLRGDVFLQPFSPVWYLFGGVLSLIGVLLWITFYMECKRMQKRWRDLPLSLLSTFLISFGAACIQSSSLIP
 RTPHNAGLRGDVFLQPFSPVWYLFGGVLSLIGVLLWITFYMECKRMQKRWRDLPLSLLSTFLISFGAACIQSSSLIP

396 RSAGGRLIYFALFLISFIMYNYTTSVVVSSLLSSPVKSKIKTMQQLAESSLTVGLEPLPFTKSYLNYSLRPEIHLFIKR 474
 RSAGGRLIYFALFLISFIMYNYTTSVVVSSLLSSPVKSKIKTMQQLAESSLTVGLEPLPFTKSYLNYSLRPEIHLFIKR
 RSAGGRLIYFALFLISFIMYNYTTSVVVSSLLSSPVKSKIKTMQQLAESSLTVGLEPLPFTKSYLNYSLRPEIHLFIKR

475 KIESQTQNPWLPAEQGVLRVRDNPYVYVFETSSGYAYVERYFTAQEIICDLNEVLFPEQLFYTHLRNSTYKELFR 553
 KIESQTQNPWLPAEQGVLRVRDNPYVYVFETSSGYAYVERYFTAQEIICDLNEVLFPEQLFYTHLRNSTYKELFR
 KIESQTQNPWLPAEQGVLRVRDNPYVYVFETSSGYAYVERYFTAQEIICDLNEVLFPEQLFYTHLRNSTYKELFR

554 LRFLRILETGYYRKQRSYVHMKLHCAQNFVITVGMEYVAPLLLMLICADILVVVILLVELAWKRFFTRPLTFHP 629
 LRFLRILETGYYRKQRSYVHMKLHCAQNFVITVGMEYVAPLLLMLICADILVVVILLVELAWKRFFTRPLTFHP
 LRFLRILETGYYRKQRSYVHMKLHCAQNFVITVGMEYVAPLLLMLICADILVVVILLVELAWKRFFTRPLTFHP

Extended Data Figure 3 | Alignment of drosophilid Ir75a orthologues. Protein-sequence alignment of *D. melanogaster*, *D. simulans* and *D. sechellia* Ir75a. Blue bars indicate the S1 and S2 lobes of the predicted LBD. The position of the PTC (X) is highlighted in yellow. Dark grey columns in the alignment highlight amino acids conserved only in two of the three

species. Pink and red shading represents *D. sechellia*-specific amino acid changes within the LBD; red denotes the subset located in the internal cavity of the binding pocket (Fig. 4a). The locations of the peptide epitopes for the Ir75a antibodies are highlighted with green dashed boxes.

1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America

Michael Worobey¹, Thomas D. Watts¹, Richard A. McKay², Marc A. Suchard³, Timothy Granade⁴, Dirk E. Teuwen⁵, Beryl A. Koblin⁶, Walid Heneine⁷, Philippe Lemey⁷ & Harold W. Jaffe⁴

The emergence of HIV-1 group M subtype B in North American men who have sex with men was a key turning point in the HIV/AIDS pandemic. Phylogenetic studies have suggested cryptic subtype B circulation in the United States (US) throughout the 1970s^{1,2} and an even older presence in the Caribbean². However, these temporal and geographical inferences, based upon partial HIV-1 genomes that postdate the recognition of AIDS in 1981, remain contentious^{3,4} and the earliest movements of the virus within the US are unknown. We serologically screened >2,000 1970s serum samples and developed a highly sensitive approach for recovering viral RNA from degraded archival samples. Here, we report eight coding-complete genomes from US serum samples from 1978–1979—eight of the nine oldest HIV-1 group M genomes to date. This early, full-genome ‘snapshot’ reveals that the US HIV-1 epidemic exhibited extensive genetic diversity in the 1970s but also provides strong evidence for its emergence from a pre-existing Caribbean epidemic. Bayesian phylogenetic analyses estimate the jump to the US at around 1970 and place the ancestral US virus in New York City with 0.99 posterior probability support, strongly suggesting this was the crucial hub of early US HIV/AIDS diversification. Logistic growth coalescent models reveal epidemic doubling times of 0.86 and 1.12 years for the US and Caribbean, respectively, suggesting rapid early expansion in each location³. Comparisons with more recent data reveal many of these insights to be unattainable without archival, full-genome sequences. We also recovered the HIV-1 genome from the individual known as ‘Patient 0’ (ref. 5) and found neither biological nor historical evidence that he was the primary case in the US or for subtype B as a whole. We discuss the genesis and persistence of this belief in the light of these evolutionary insights.

No comprehensive genomic analysis of the emergence and early spread of HIV-1 in North America—where HIV/AIDS was first recognized—has been possible because the only pre-1980 HIV-1 group M genome currently available (strain Z321B) was sampled in Africa. To fill this gap, we performed serological screening and viral genome sequencing of archived serum samples dating back to 1978–1979, originally collected from men who have sex with men (MSM) cohort patients in New York City (NYC) and San Francisco (SF). NYC samples were from volunteers in a prospective study of AIDS established in 1984 (ref. 6), 378 of whom had been part of an earlier cohort of 8,906 men involved in hepatitis B virus (HBV) studies beginning in 1978 (ref. 7), and for which stored sera from 1978 and/or 1979 were available⁸. Previous work showed that 6.6% of these sera from NYC in 1978–1979 were HIV-1 seropositive⁶; 33 of these positive samples were chosen for attempted HIV-1 sequencing. The sera from SF originated from a study of approximately 6,875 patients enrolled in the late 1970s in HBV studies at the San Francisco City Clinic⁹. We tested 2,231 of these samples from 1978 and found, by western blot, that 83 (3.7%) were

positive for HIV-1 antibodies; of these, 20 were randomly chosen for attempted HIV-1 sequencing.

Low template number and degradation arising from long-term storage were major challenges for genomic analysis, as encountered previously with similar samples¹⁰: recovered RNA was generally below the limits of quantification and initial attempts at amplification of reverse-transcribed viral RNA failed consistently and indicated that viral RNA survived in the 1970s samples only in short fragments. This led us to design an RNA ‘jackhammering’ approach to greatly increase both the ability to detect viral RNA-positive samples and to recover complete genomic HIV-1 sequences from them. Briefly, we used large panels of primers to amplify many short fragments in separate pools, such that amplicons overlapped between but not within each pool (Extended Data Fig. 1 and Supplementary Table 1). Each pool’s amplicon set filled gaps between those of complementary pools, with the entire panel providing complete genomic coverage. Moreover, a preliminary, multiplex amplification step greatly concentrated target RNA before final amplification and sequencing.

Three samples from SF and five from NYC provided sufficient data to assemble coding-complete sequences. Bayesian phylogenetic analyses of these HIV-1 genomes (Fig. 1 and Extended Data Fig. 2) showed that although they were the oldest sampled outside Africa, they do not fall on the deepest branches even within subtype B. Instead, the 1970s genomes and the US epidemic as a whole were phylogenetically nested within the more genetically diverse, older subtype B epidemic in Caribbean countries. Separate analyses of *gag*, *pol* and *env* sequences also placed the US sequences in a strongly supported monophyletic clade nested within the paraphyletic Caribbean subtype B sequences from Haiti, Dominican Republic, Jamaica, Trinidad and Tobago and Haitian immigrants in the US (Extended Data Figs 3, 4). Molecular clock phylogeographic analysis of the complete genome data supported a subtype B ancestor in the Caribbean (posterior probability >0.99) dating to 1967 (95% credibility interval 1963–1970) (Extended Data Table 1). This provided genome-wide evidence that the epidemic moved from the Caribbean to the US rather than from the US to the Caribbean².

Location transition estimates recovered a relatively precise date (1971 (1969–1973), Extended Data Table 1) for the HIV-1 jump from the Caribbean, very shortly before the US most recent common ancestor (MRCA). This narrow timing is aided by the basal relationship of a very close relative from the Caribbean (sequence ‘H6’ from an individual who entered the US from Haiti in 1981)² (Extended Data Fig. 2). The probability density of the date of introduction to the US overlaps with the deep branching structure in Caribbean diversity (Fig. 1 and Extended Data Fig. 3), indicating that the US clade emerged from the Caribbean epidemic during its early growth phase. We estimated a relatively fast logistic growth rate of 0.62 (0.26–0.99) yr⁻¹ within the Caribbean population (Fig. 2). That of the US population was even

¹Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA. ²Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, UK. ³Departments of Biomathematics, Biostatistics and Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. ⁴Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA. ⁵UCB, Brussels BE-1070, Belgium. ⁶Laboratory of Infectious Disease Prevention, The New York Blood Center, New York, New York 10065, USA. ⁷Department of Microbiology and Immunology, Rega Institute, KU Leuven—University of Leuven, Minderbroedersstraat 10, 3000 Leuven, Belgium.

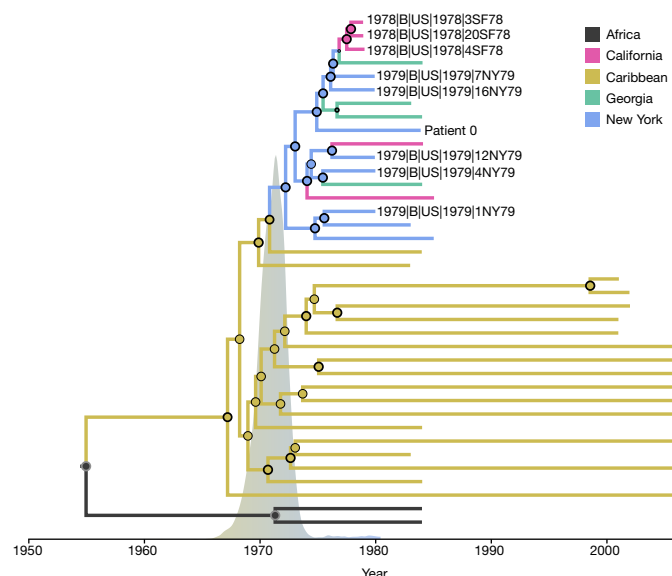


Figure 1 | Maximum clade credibility (MCC) tree summary of the Bayesian spatio-temporal reconstruction based on complete HIV-1 genome data. The tips of the tree correspond to year of sampling while branch and node colours reflect the sampling locations for the tip branches and the inferred locations for the internal branches. Tip labels are provided for the newly obtained archival HIV-1 genomes. Diameters of internal node circles reflect posterior location probability values. Thick outer circles represent internal nodes with posterior probability support >0.95 . We also depict the posterior probability density (grey) for the time of the introduction event from the Caribbean into the US on the time scale of the tree. A fully annotated tree for this data set ('full genome 38', which includes only sequences sampled early in the US epidemic) is shown in Extended Data Fig. 2b; 'full genome 46' which includes all available complete genomes basal to the 'pandemic clade'² of subtype B, plus a similar number and date range of US pandemic clade sequences, is shown in Extended Data Fig. 2a. Separate analyses of *gag*, *pol*, *env*, and the coding-complete genomes (including also sequences sampled later in the US epidemic) provide consistent results (Extended Data Figs 3, 4).

higher, 0.81 (0.65 – 0.98) yr^{-1} , in line with a precipitous spread among existing high-risk sexual networks. These mean growth rate estimates corresponded to doubling times of 1.12 years and 0.86 years for the Caribbean and the US, respectively; both the more rapid and longer growth in the US appear to have contributed a higher number of 'effective infections' (Fig. 2), with the US overtaking the Caribbean by around 1977 despite the later HIV-1 emergence in the US.

Molecular clock analyses of larger numbers of *env* sequences revealed similar time of the most recent common ancestor (TMRCA) estimates for the key nodes (Fig. 3, Extended Data Table 1 and Extended Data Figs 5, 6). Interestingly, our modest snapshot of 1970s sequences from NYC and SF (Fig. 3, Extended Data Fig. 5b) encompassed the full diversity exhibited by HIV-1 sequences from later years (that is, it shares the same MRCA as larger sequence sets sampled in later years): all post-1985 US sequences are nested within the early diversity captured by the limited number of 1970s sequences we recovered (Extended Data Fig. 6).

A phylogeographic reconstruction including only those US sequences sampled from known locations between 1978–1984 (Fig. 1) demonstrated that the NYC epidemic was already relatively mature and genetically diverse by 1979, tracing back to an MRCA estimated at 1972 (1970–1974) and there is strong support for the idea that the US subtype B ancestor circulated in NYC (posterior probability = 0.99). Indeed, the extensive genetic diversity in the US (and in NYC in particular) in 1978–1979 can be explained only by several years of circulation of the virus before 1978–1979.

Using sequences sampled from NYC, North Carolina and California relatively late in the epidemic (comparable to the 1978–1984 East coast, West coast and Southern sampling), we still inferred a US ancestor in

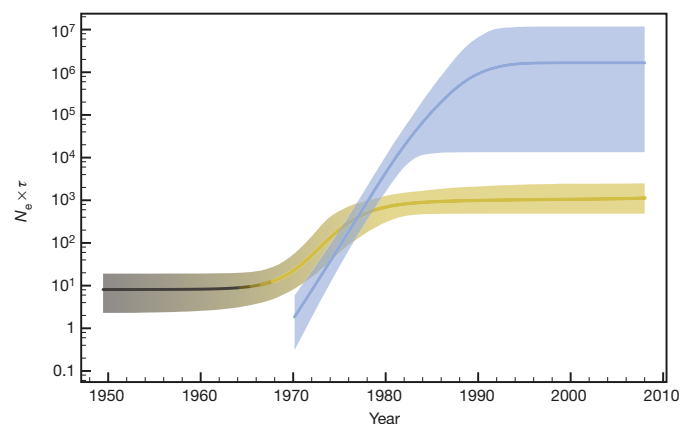


Figure 2 | Demographic reconstruction based on the nested coalescent model. The colour scheme is consistent with that of the phylogeographic analyses in Figs 1 and 3: the constant-logistic population size estimates (the 'effective number of infections', N_e , multiplied by the mean viral generation time, τ) through time are depicted in a black–yellow colour range (following the African and Caribbean locations in the phylogeographic analyses) while the logistic population size estimates for the nested US clade are shown in blue (as for the US/NYC location in the phylogeographic analyses).

NYC, but with only modest support that prevents us from drawing firm conclusions ($pp = 0.67$, Extended Data Figure 6b and Extended Data Table 1). As a generality, early samples close to the deep branching structure are essential to confidently reconstruct the initial spatio-temporal expansion dynamics in exponentially growing populations.

Compared to NYC, the SF epidemic in 1978 appeared to have been established more recently (Figs 1, 3, and Extended Data Figs 2b, 5b). It is striking that all three independently detected complete HIV-1 genomes we found are so closely related; moreover, they form a cluster with three partial *env* sequences sampled in SF during the same period¹⁰ (Extended Data Fig. 5b). This suggests that the bulk of the HIV-1 infections in SF in 1978 traced back to a single introduction from NYC in around 1976 (consistent with the lower HIV-1 seroprevalence in the SF cohort).

The sampled sequences thus reveal a series of key founder events in the genesis of subtype B (for example, Fig. 3 and Extended Data Table 1), with the epidemic spreading from the African HIV-1 group M epicentre to the Caribbean by about 1967, from the Caribbean to NYC by about 1971 and from NYC to SF by about 1976, quickly followed by extensive geographical mixing in the US and beyond.

Reports of one cluster of homosexual men with AIDS linked through sexual contact were important in suggesting the sexual transmission route of an infectious agent before the identification of HIV-1 (refs 5, 11). Beginning in California, CDC investigators eventually connected 40 men in ten American cities to this sexual network. Investigators placed one man with Kaposi's sarcoma near the centre of a sociogram representing this cluster and identified him as 'Patient 0'—a 'non-Californian AIDS patient' and a possible 'carrier' of an infectious agent (Extended Data Fig. 7). Before publication, Patient 'O' was the abbreviation used to indicate that this patient with Kaposi's sarcoma resided 'Out(side)-of-California.' As investigators numbered the cluster cases by date of symptom onset, the letter 'O' was misinterpreted as the number '0', and the non-Californian AIDS patient entered the literature with that title¹². Although the authors of the cluster study repeatedly maintained that Patient 0 was probably not the 'source' of AIDS for the cluster or the wider US epidemic, many people have subsequently employed the term 'patient zero' to denote an original or primary case, and many still believe the story today¹³. We therefore recovered the complete HIV-1 genome of Patient 0 and examined it against the backdrop provided by the 1970s sequences.

Although labelled as the cluster study's 'index patient', Patient 0 was neither the first AIDS case to come to CDC researchers'

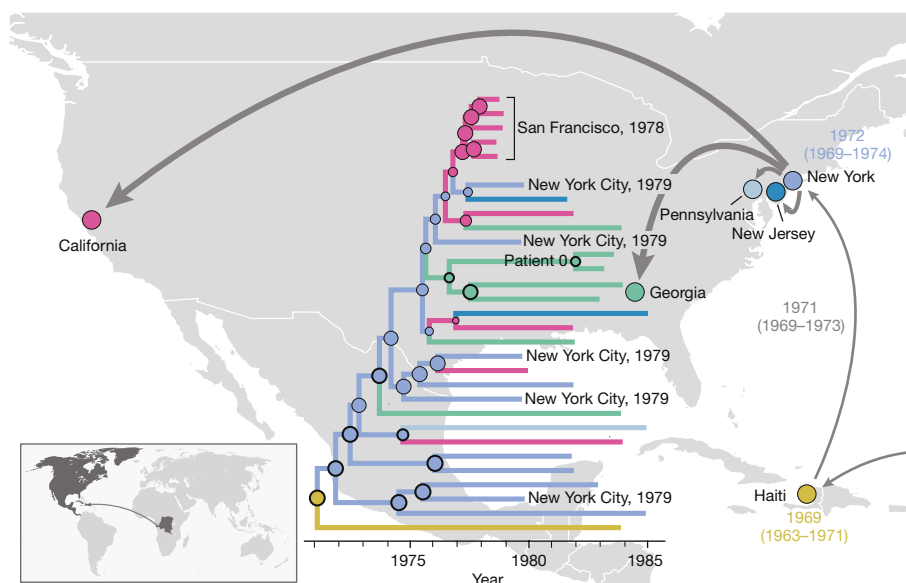


Figure 3 | The early patterns of HIV-1 subtype B spread in the Americas. The map summarizes the main patterns of spread inferred from the molecular clock phylogeographic analyses. The map inset shows the initial introduction of the subtype B lineage into the Caribbean from Africa. From there, the virus spreads first to NY and subsequently to different locations in the United States. The tree depicts the US clade, plus the most closely related basal Haiti strain, as inferred from the ‘*env* 74’ analysis (Extended Data Fig. 5b). Tips of the clade correspond to the year of sampling. Tip branch colours reflect the actual sampling locations as indicated on the map; interior branches depict phylogenetically inferred locations using the same colour scheme. Diameters of internal node circles

reflect posterior location probability values. Thick outer circles indicate internal nodes with posterior probability support >0.95 . Thickness of the arrows reflects number of transitions inferred from this tree cluster. Mean dates and 95% credible intervals in yellow and blue represent the date estimates for the MRCA in the Caribbean and the US, respectively, based on the *env* 74 analysis. Date next to arrow between these locations represents the estimated timing of the corresponding jump. Patient 0 (represented by two sequences) and the earliest sequences from San Francisco (1978) and New York City (1979) are labelled. Maps made with Natural Earth.

attention, nor the first to display symptoms. In general, the CDC numbered cases in the order that the reports reached the agency from different cities and employed the terms ‘cases’ and ‘patients’ interchangeably. Patient 0, until he was linked to the cluster and took on his new name, was Case (or Patient) 057. The cluster study’s LA 6 was the CDC’s Case 032, and several cases in the New York section of the cluster⁵ (Extended Data Fig. 7) were also reported before Patient 0 (and thus brought to investigators’ attention first): NY 3 was Case 001, NY 2 was Case 002, NY 6 was Case 010 and NY 5 was Case 053 (ref. 14).

The information available for CDC investigators to establish symptom onset dates was often fragmentary and thus resisted uniform categorization. Sometimes onset was determined on the basis of lymphadenopathy, other times by the appearance or diagnosis of Kaposi’s sarcoma or *Pneumocystis carinii* pneumonia. Investigators were unable to link to the cluster several NYC-based cases that had much earlier dates of symptom onset. For example, Case 154 was a middle-aged European man whose reported onset date for Kaposi’s sarcoma was January 1975; and Case 153, when he was diagnosed with Kaposi’s sarcoma in September 1981, recalled having swollen glands as early as June 1977 (ref. 15). Yet even within the cluster, Case 057’s symptoms (lymphadenopathy in December 1979 and a Kaposi’s sarcoma lesion diagnosed in May 1980, ref. 5) appeared considerably later than those of several other cases. LA 1 (Case 335) developed a lesion in February 1978 (ref. 16), whereas NY 1 (Case 152) experienced the onset of Kaposi’s sarcoma in December 1978, NY 2 (Case 002) in May 1979 and NY 3 (Case 001) in August 1979 (ref. 14).

In his book *And the Band Played On*, Randy Shilts identified ‘Patient Zero’ by name as a highly sexually active French-Canadian flight attendant¹⁷. Unlike the initial reports of the cluster, media coverage of Shilts’s book strongly insinuated that this individual was the source of the North American epidemic and an exemplar of dangerous disease transmission¹⁸—ideas which found a global audience (Supplementary

Discussion). However, we found that the HIV-1 genome from this individual appeared typical of US strains of the time and was not basal to the US diversity, let alone to the deeper Caribbean subtype B diversity, in a manner that might be suggestive of a special role (Figs 1, 3). In short, we found no evidence that Patient 0 was the first person infected by this lineage of HIV-1.

In addition to donating plasma for analysis, Patient 0 provided investigators with the names of nearly 10% of his sexual partners over several years⁵, while many other cluster patients were unable to share more than a handful of names¹⁶. This strongly suggests that ascertainment bias contributed to his central role in the cluster study and its diagrammatic representation. Later research would also call into question the cluster study’s estimated average latency period of 10.5 months between sexual contact and symptom onset, with a revised average incubation period approaching 10 years for MSM. In retrospect, the study’s sociogram (Extended Data Fig. 7) almost certainly depicted the sexual contacts of these men years after they had contracted HIV-1 (ref. 19) (Supplementary Discussion). Other East coast HIV-1 sequences fall much closer to the main early-California clade we identify than does that of Patient 0 (Fig. 3). Thus, while he did link AIDS cases in New York and Los Angeles through sexual contact, our results refute the widespread misinterpretation that he also infected them with HIV-1.

Much like historical reconstructions, phylogenetic inferences are often generated from data collected long after the critical events occurred. Our work highlights the importance of complete viral genomes from early archival specimens, carefully contextualized through historical analysis, without which this detailed picture of these early landmarks in the HIV/AIDS pandemic would not have been possible.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 April; accepted 7 September 2016.

Published online 26 October 2016.

1. Korber, B. *et al.* Timing the ancestor of the HIV-1 pandemic strains. *Science* **288**, 1789–1796 (2000).
2. Gilbert, M. T. *et al.* The emergence of HIV/AIDS in the Americas and beyond. *Proc. Natl Acad. Sci. USA* **104**, 18566–18570 (2007).
3. Holmes, E. C. When HIV spread afar. *Proc. Natl Acad. Sci. USA* **104**, 18351–18352 (2007).
4. Pape, J. W. *et al.* The epidemiology of AIDS in Haiti refutes the claims of Gilbert *et al.* *Proc. Natl Acad. Sci. USA* **105**, E13 (2008).
5. Auerbach, D. M., Darrow, W. W., Jaffe, H. W. & Curran, J. W. Cluster of cases of the acquired immune deficiency syndrome. Patients linked by sexual contact. *Am. J. Med.* **76**, 487–492 (1984).
6. Stevens, C. E. *et al.* Human T-cell lymphotropic virus type III infection in a cohort of homosexual men in New York City. *J. Am. Med. Assoc.* **255**, 2167–2172 (1986).
7. Szmuness, W., Stevens, C. E., Zang, E. A., Harley, E. J. & Kellner, A. A controlled clinical trial of the efficacy of the hepatitis B vaccine (Heptavax B): a final report. *Hepatology* **1**, 377–385 (1981).
8. Koblin, B. A., Morrison, J. M., Taylor, P. E., Stoneburner, R. L. & Stevens, C. E. Mortality trends in a cohort of homosexual men in New York City, 1978–1988. *Am. J. Epidemiol.* **136**, 646–656 (1992).
9. Jaffe, H. W. *et al.* The acquired immunodeficiency syndrome in a cohort of homosexual men. A six-year follow-up study. *Ann. Intern. Med.* **103**, 210–214 (1985).
10. Foley, B., Pan, H., Buchbinder, S. & Delwart, E. L. Apparent founder effect during the early years of the San Francisco HIV type 1 epidemic (1978–1979). *AIDS Res. Hum. Retroviruses* **16**, 1463–1469 (2000).
11. Centers for Disease Control (CDC) A cluster of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. *MMWR Morb. Mortal. Wkly. Rep.* **31**, 305–307 (1982).
12. McKay, R. A. *Imagining 'Patient Zero': Sexuality, Blame, and the Origins of the North American AIDS Epidemic*. Doctoral thesis, Univ. of Oxford (2011).
13. Harden, V. A. *AIDS at 30: A History* (Potomac Books, 2012).
14. Darrow, W. W. Trip report to New York City, July 12–16 and August 3–6, 1982. CDC Task Force on AIDS, internal communication (3 September 1982).
15. Darrow, W. W. Time-space clustering of KS cases in the City of New York: evidence for horizontal transmission of some mysterious microbe. CDC Task Force on Kaposi's Sarcoma and Opportunistic Infections, internal communication (3 March 1982).
16. Darrow, W. W. & Auerbach, D. M. Los Angeles cluster: background. CDC Task Force on Kaposi's Sarcoma and Opportunistic Infections, internal communication (12 May 1982).
17. Shilts, R. *And the Band Played On: Politics, People, and the AIDS Epidemic* (St. Martin's Press, 1987).
18. McKay, R. A. "Patient Zero": the absence of a patient's view of the early North American AIDS epidemic. *Bull. Hist. Med.* **88**, 161–194 (2014).
19. Moss, A. R. In response to: AIDS without end. *New York Rev. Books* **35**, 60 (1988).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Stevens and D. Hemmerlein for facilitating access to archival sera; G.-Z. Han, A. Bjork, W. Switzer, V. Sullivan, R. Ruboyianes and P. Sprinkle for technical assistance; T. Spira and M. Owen for geographical data on some published sequences; and the NIH AIDS Reagent program for providing reference virus samples US657 and HT599. W. W. Darrow led the initial 1982 cluster investigation and provided R.A.M. with access to his copies of archival CDC documents. This work was supported by NIH/NIAD R01AI084691 and the David and Lucile Packard Foundation (M.W.); the Wellcome Trust (080651), the University of Oxford's Clarendon Fund, the Economic and Social Research Council (PTA-026-27-2838), and a J. Armand Bombardier Internationalist Fellowship (R.A.M.); the Research Fund KU Leuven (Onderzoeksfonds KU Leuven, Program Financing no. PF/10/018) and the 'Fonds voor Wetenschappelijk Onderzoek Vlaanderen' (FWO) (G066215N) (P.L.); and NSF DMS 1264153, NIH R01 HG006139 and NIH R01 AI107034 (M.A.S.).

Author Contributions M.W., H.W.J., P.L. and R.A.M. conceived the study. T.D.W. and M.W. designed the RNA jackhammering method. T.D.W. generated the sequences. B.A.K. provided serum samples from New York City. W.H. and T.G. acquired specimens and provided serological data. D.E.T. provided conceptual input. M.W., M.A.S. and P.L. prepared the data sets and performed the phylogenetic analyses. R.A.M. performed the historical analyses. M.W., H.W.J., P.L. and R.A.M. wrote the paper. All authors discussed the results and commented on the manuscript. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W. (worobey@email.arizona.edu) or R.A.M. (ram78@cam.ac.uk).

Reviewer Information *Nature* thanks K. Andersen and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

HIV-1 serological screening of serum samples from San Francisco from 1978. We tested 2,231 samples collected from the MSM cohort in San Francisco in 1978 (ref. 9) and detected 83 HIV-1-positives by Western Blot (3.7% prevalence). Samples were first screened by GS HIV-1/HIV-2 Plus O EIA (Bio-Rad Laboratories) and reactive samples were further tested by WB Genetic Systems HIV-1 Western Blot (Bio-Rad Laboratories).

HIV-1 nucleic acid amplification. A total of 33 samples of frozen serum from New York City previously identified as positive for antibody to HIV-1^{6–8} were assayed; and a total of 20 frozen serum samples from San Francisco⁹, identified as part of the present study as positive for antibody to HIV-1, were assayed. The New York City samples were from 1978 and 1979 though no complete genomic sequences from 1978 were developed. The San Francisco samples were all from 1978. RNA recovered from samples from both NY and SF was generally undetectable when assaying 5- μ l aliquots in a Qubit 2.0 fluorometer using the Qubit RNA HS reagents (detection limit, 250 pg μ l⁻¹).

Additionally, a sample of peripheral blood mononuclear cells (PBMCs) and a sample of serum were both assayed; these had been collected from a single individual in 1983 (Patient 0), and the samples were stored at CDC Atlanta. Other than Patient 0, now deceased, the data recorded were unlinked to individual identifiers and the work was approved by the Human Subjects Protection Program at the University of Arizona.

Four panels of degenerate primers (Supplementary Table 1 and Extended Data Fig. 1) were designed using a suite of North American subtype B sequences. We aimed to design primers able to amplify both conserved regions and predictably variable sites. Primers within each panel were designed to generate sequence from the 5' end of *gag* to the 3' end of *nef* and were designed to amplify overlapping fragments. Two panels 'HIVL' ($n = 25$) and 'HIVLb' ($n = 22$) were designed to amplify fragments of approximately 500–650 bases in length. Two other panels 'HIVM' ($n = 50$) and 'HIVR' ($n = 46$) were designed to amplify fragments of approximately 200–320 bases in length.

Nucleic acids from 100- μ l aliquots of serum (or PBMCs in the case of Patient 0) were isolated using the QIAamp viral RNA mini kit (Qiagen) with 5 mcg added carrier RNA. Serum samples were then treated with DNase I (Invitrogen, Life Technologies) before reverse transcription. PBMC nucleic acids were left untreated.

Proximal DNA from Patient 0's PBMCs was amplified with all four primer panels and from multiple separate isolations. Amplification was achieved using Invitrogen platinum Taq DNA polymerase high fidelity (Life Technologies) and run for 55 cycles at an annealing temperature of 52 °C. Additionally, attempts were made to amplify longer fragments using PCR supermix high fidelity (Life Technologies) and forward and reverse primers matched from the HIVLb primer panel for long fragment length followed by nesting with primers for slightly shorter fragment length. A single fragment of slightly more than 7,000 bases was generated after multiple attempts with multiple primer combinations and cloned using the Invitrogen TOPO XL PCR cloning kit (Life Technologies). Fragments of individual clones were then amplified using HIVLb forward and reverse primers matched to give approximately 1,000-base overlapping fragments and then sequenced.

RNA jackhammering. RNA jackhammering of the serum samples proceeded as follows: aliquots of RNA extract were reverse transcribed using the GoScript reverse transcription system (Promega) using a program of 4 cycles of 50 °C for 30 min followed by 55 °C for 30 min and a final incubation at 85 °C for 10 min. Primers used were pools of reverse primers from widely spaced amplicons (Supplementary Table 1, Extended Data Fig. 1), typically nine or ten primers per pool in a single reaction tube, with the wide spacing abrogating the possibility of incorporation of an internal primer into any given amplicon. Reverse transcription products were then briefly amplified in multiplex reactions in the pool-specific tube (denaturation for 3 min at 94 °C followed by 30 cycles of 94 °C for 30 s, 52 °C for 30 s, 68 °C for 30 s, and a final extension of 68 °C for 5 min) with matching forward primer pools (a 'preliminary amplification' step). Sequences were then amplified from individual aliquots taken from the pool-specific tubes, via single primer pairs (denaturation for 3 min at 94 °C followed by 40 cycles of 94 °C for 30 s, 52 °C for 30 s, 68 °C for 30 s, and a final extension of 68 °C for 5 min). Two separate isolates were amplified from each sample in this manner, with a minimum of one amplification with each primer panel per isolate. Five out of the 33 (15%) of the NY sera assayed yielded complete HIV-1 genomic data as did 3 out of the 20 (15%) SF sera, suggesting that levels of viral RNA preservation were very similar in each collection.

In Extended Data Fig. 1 we schematically illustrate the RNA jackhammering approach and its advantages over standard RT-PCR procedures for degraded, low input samples. For a conventional RT-PCR approach with a fairly long amplification

product we would perform reverse transcription and obtain one potentially amplifiable cDNA product. We would then aliquot ~10% of the reverse transcription product for amplification in a PCR reaction with forward and reverse primers. Even if the single cDNA product made it into the PCR reaction, the desired amplification product would be too long and a PCR amplicon would therefore not be obtained. For RT-PCR with a shorter amplification product, more appropriately sized given the damaged RNA in the sample, there was still a 90% chance that it would be deemed a negative sample since most aliquots will not contain the rare cDNA product. Using multiple primer sets would increase the chance of a PCR-positive result, but most PCR reactions remained negative because most aliquots lack target cDNA. Even with a 10 primer-pair pool and 10 final PCR reactions, there may be no amplified product. The RNA jackhammering approach targets large panels of appropriately short amplicons, uses discrete pools of non-overlapping primers pairs for reverse transcription, and includes a crucial multiplex pre-amplification step to ensure that each aliquot contained ample template molecules for the final PCR amplification (a separate reaction for each primer pair in the entire panel).

Sequencing was performed at the University of Arizona Genetics Core using an ABI 3730XL. The Patient 0 sample contained considerable heterogeneity (mixed bases) both in proviral assembly and in viral RNA amplification. Heterogeneity in the NY and SF samples (all sequences derived from viral RNA) was low. In all cases consensus sequences were used in the phylogenetic analyses. Primer sequences were computationally removed from all sequence data before assembling genomic consensus sequences, which yielded coding-complete genomic data with exception of a few small gaps and the 3' end of the *nef* gene (Supplementary Table 2).

Validation of the jackhammering approach. To validate this approach we obtained seed stock samples from the NIH AIDS Reagent program of subtype B viruses from the US (US657) and Haiti (HT599) and applied a jackhammering approach with independent runs of both the HIVM and HIVR primer panels (Extended Data Fig. 8).

For US657 we recovered, in total, from both runs combined, 8,194 nt of high quality data. HIVM and HIVR are independent runs with completely different primer sets, yet where the data overlapped, they were >99.9% similar. Moreover, the few heterogeneities did not line up with heterogeneous primers but fell in regions between primers, demonstrating that differences could not be attributed to the incorporation of primers into the recovered sequences. This was expected both because the wide spacing of amplicons within a single pool of primer pairs prevents incorporation of primers within amplified products and because all primer sequences from final amplification products were computationally removed from the sequences before assembly of genomic sequences. There are 3,354 bases in the published US657 sequence. Our data covered about 90% of the 3,354 bases of previously published US657 sequence (GenBank accession number U04908) and all of our individual amplicons in the region of overlap had US657 as the highest BLAST hit and were >99% similar to the published sequence.

For HT599 the HIVM and HIVR primer panels developed 8,545 nt of data, 99.6% of the target. The HIVM-derived sequence was >99.9% similar to the HIVR-derived sequence. We recovered 100% of the overlap with the previously published HT599 sequence (2,881 nt, GenBank accession number U08447) with 99.5% similarity.

To evaluate discrepancies between the jackhammering-recovered sequences and both US657 and HT599, we compared consensus sequences of combined HIVM and HIVR data with the respective published sequences by adding them to our complete genome alignment and reconstructing a maximum likelihood tree (Extended Data Fig. 8a). As expected, the independently generated sequences from each virus clustered very closely and only had short tips from their common ancestors, resulting from a very small number of substitutions in their overlapping regions. In a root-to-tip analysis (Extended Data Fig. 8b), our sequences (with a target symbol) were associated with somewhat smaller residuals than the published sequences (with a circle), indicating that our data are likely to be more accurate and, importantly, cannot contain primer remnants as this would result in much larger residuals.

Sequence data. To construct the data sets for the analyses shown in Fig. 1 and Extended Data Figs 2–4 we searched the Los Alamos National Laboratories (LANL) HIV database (<http://hiv.lanl.gov/>) for all available genome-length HIV-1 sequences from Caribbean countries, which had previously been shown to exhibit diverse subtype B lineages that fall basal to a monophyletic 'pandemic' clade of subtype B that accounts for most US and other non-Caribbean subtype infections². These included sequences sampled in Haiti, Dominican Republic, Jamaica and from Haitians who had recently immigrated to the US from Haiti ('H3' and 'H5' from 1982, 'H6' and 'H7' from 1983, 'RF_HAT' from 1983)². For sequences H3, H5, H6 and H7 *pol* sequences were not available, but partial *gag* and full-length *env* sequences were available. For the full-genome analyses the *pol* gene was treated as missing data. We then added a similar number of genomes from the US from

a similar time period (1982–2005), plus one each from France and the UK, as well as outgroup sequences of subtype D from the Democratic Republic of the Congo (D.R.C.). We called this the ‘full genome 46’ data set because it contained 46 genomes. The *gag*, *pol* and *env* data sets depicted in Extended Data Fig. 3 were each derived from the respective sub-genomic region of this same set of taxa. The subset of ‘full genome 46’ that contained only those US sequences sampled from 1978–1984 we called ‘full genome 38’.

For the *env* analyses in Fig. 3 and Extended Data Fig. 5 the alignment from ref. 2 was used, with the addition of the sequences generated for the present study, additional Caribbean subtype B sequences from 2000 to 2005, and four early subtype B partial *env* sequences from San Francisco¹⁰. This alignment we called ‘*env* 105’. The subset that contained only those US sequences sampled from 1978–1984 we called ‘*env* 74’.

For Extended Data Fig. 6 we added to ‘*env* 105’ a comparable number—relative to those sampled from 1978–1984 from known locations (New York, California, Georgia, Pennsylvania, New Jersey) (Extended Data Fig. 4b)—of randomly sampled sequences from 1997–2007 from NY, SE and North Carolina (the closest available site with sufficient numbers to stand in for the Georgia ones from the 1978–1984 sample). We called this alignment ‘*env* 133’.

In all cases sequences were manually aligned using Se-AL (<http://tree.bio.ed.ac.uk/software/seal/>). All sequence alignments, input files, tree files and primer sequences are available at the Dryad Digital Repository (doi:10.5061/dryad.7mv7v).

Recombination analysis and maximum likelihood tree reconstruction. Maximum likelihood phylogenies were reconstructed using RAXML under on a general time-reversible model of substitution with gamma distributed rate variation among sites²⁰. Bootstrap support values were calculated using 1,000 pseudo-replicates. To detect the presence of recombination, we first performed the Phi test²¹ on every data set (Extended Data Table 1). When the null hypothesis of absence of recombination was rejected ($P < 0.05$), we subsequently analysed the data set using RDP4 (ref. 22) and produced new alignments in which the minor recombinant regions were deleted from putative recombinants. Re-analyses of these ‘recombination-free’ data sets using the Phi test confirmed the absence of detectable recombination signal ($P > 0.05$, Extended Data Table 1).

Bayesian phylogenetic inference. Time-measured phylogeographic histories were reconstructed using a Bayesian phylogenetic inference approach implemented in BEASTv1.8.2 (ref. 23). Our full probabilistic model combined sequence substitution over an unknown phylogeny calibrated in time units using a molecular clock process with dated tips²⁴, a coalescent tree prior and a discrete diffusion process among discrete location states²⁵. For the sequence substitution process, we used the same model as for the maximum likelihood reconstructions. We accommodated rate variation among lineages using a lognormal distribution in an uncorrelated relaxed molecular clock model²⁶ and integrated out each sampling date over an uncertainty interval of one year. Visual inspections of root to tip divergence as a function of sampling time using TempEst²⁷ indicated a strong temporal signal with no clear outlier sequences (Extended Data Fig. 9).

For most analyses, we flexibly modelled changes in effective population size through time by specifying a Bayesian skygrid non-parametric tree prior with a grid of 50 years and yearly effective population size parameters²⁸. (The notion of ‘effective population size’, or ‘effective infections’ in epidemiological applications, comes from population genetics, and is typically lower than the full (that is, census) population size, reflecting, for example, variance in reproductive success among individuals—transmissions to new hosts in this context). To estimate viral population growth rates in both the Caribbean and US populations, we fitted a ‘nested’ coalescent model to the data set with the largest taxon sampling (*env* 133). This model fits a constant-logistic demographic function²⁹ to the genealogy excluding the US clade. The initial constant phase was included in the model to accommodate the deep branching between the subtype B sequences and the African subtype D outgroup sequences. Nested within this model, a separate logistic growth model was fitted to the US clade in the genealogy.

The process of discrete diffusion among locations was modelled using a general non-reversible substitution model³⁰. In our analyses including the African subtype D outgroup lineages, we set the root state frequency to one for the African state and zero for all other possible discrete states. We obtained estimates of the transitions among locations (Markov jumps) using a stochastic mapping implementation capable of inferring the complete Markov jump history^{31,32}. We approximate the posterior distribution for our full probabilistic model using Markov chain Monte

Carlo (MCMC) sampling. We use BEAGLE in conjunction with BEAST to improve the computational performance of our analyses³³. MCMC chains were run for 50,000,000 generations, sampling every 5,000 generations. We diagnosed the runs by examining trace plots and effective sample sizes, and summarized continuous parameters (mean and 95% highest posterior density (HPD) intervals) using Tracer (<http://tree.bio.ed.ac.uk/software/tracer/>) after discarding a 10% burn-in. Trees were summarized as maximum clade credibility trees using TreeAnnotator and visualized in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

In two specific phylogeographic analyses, we assessed (i) to what extent sequences sampled early in the US epidemic characterize the subtype B diversity in the US clade (Extended Data Fig. 6a) and (ii) to what extent the location state at the origin of the US clade can be estimated using sequences sampled later in the epidemic from three different US states (Extended Data Fig. 6b). For this purpose, we first reconstructed time-measured phylogenies for the *env* 133 data set using the substitution model, molecular clock model and coalescent model described above and subsequently reconstructed ancestral locations on the inferred posterior distribution of trees.

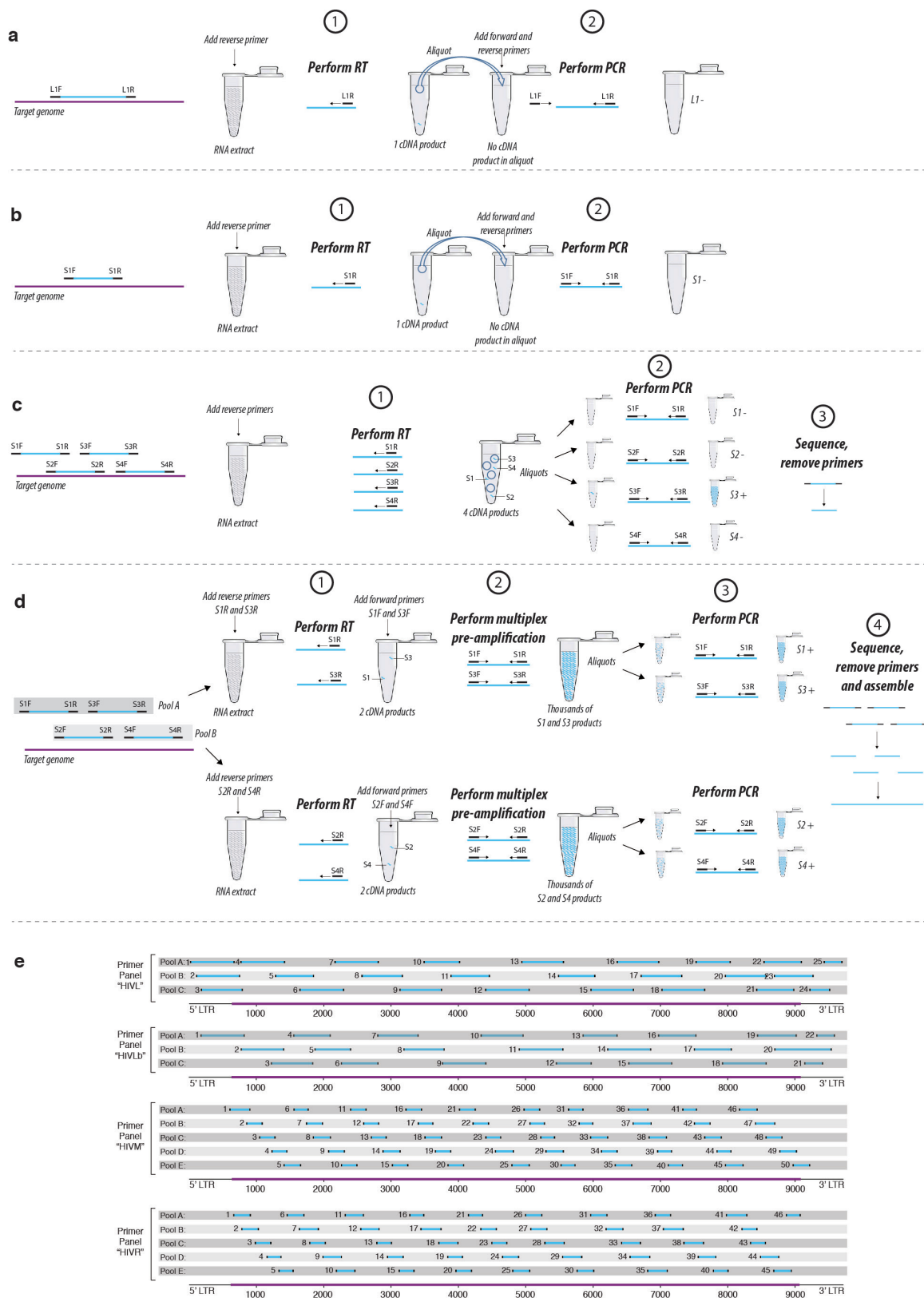
For Extended Data Fig. 6a, we classified US sequences as ‘early’ or ‘late’ depending on whether they were sampled before or after (and including) 1985. For Extended Data Fig. 6b, we first pruned the necessary US sequences from the posterior distributions in order to retain only ‘late’ sequences from New York, North Carolina and California (matching the sampling from New York, Georgia and California in Fig. 3 and Extended Data Fig. 5b). In this case, the support for a NYC ancestral state is likely upheld by the presence of two basal NYC representatives, but location estimates in a star-like tree structure with long tip branches will be critically dependent on how well the diversity of any location is represented in the contemporaneous sampling, as recently noted³⁴.

Comparison of phylogeographic estimates before and after deleting minor recombinant regions from putative recombinants (Extended Data Table 1) indicated highly consistent results.

Data availability. All sequence alignments, input files, tree files and primer sequences are available at the Dryad Digital Repository (doi:10.5061/dryad.7mv7v).

The HIV-1 sequences reported here have been deposited in GenBank under accession numbers KJ704787, KJ704788, KJ704789, KJ704790, KJ704791, KJ704792, KJ704793, KJ704794, KJ704795, KJ704796 and KJ704797.

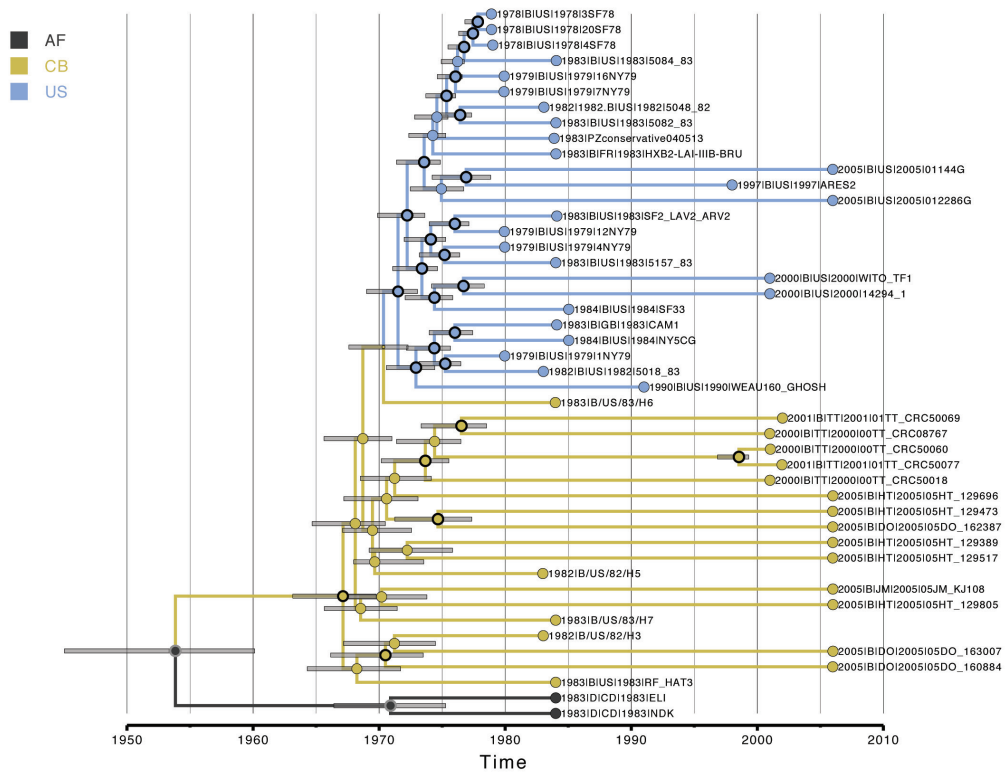
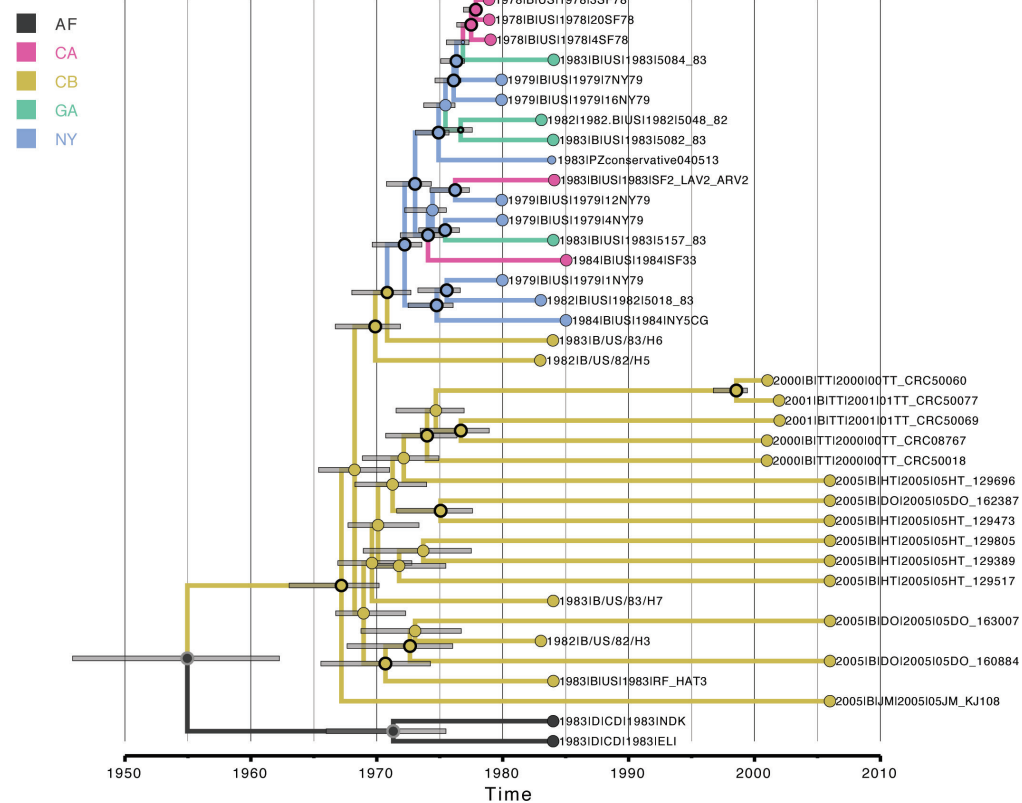
20. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
21. Bruen, T. C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681 (2006).
22. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, vev003 (2015).
23. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
24. Rambaut, A. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16**, 395–399 (2000).
25. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLOS Comput. Biol.* **5**, e1000520 (2009).
26. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
27. Rambaut, A., Lam, T. T., de Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst. *Virus Evol.* **2**, DOI: <http://dx.doi.org/10.1093/ve/vev007> (2016).
28. Gill, M. S. et al. Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
29. Faria, N. R. et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014).
30. Edwards, C. J. et al. Ancient hybridization and an Irish origin for the modern polar bear matriline. *Curr. Biol.* **21**, 1251–1258 (2011).
31. Minin, V. N. & Suchard, M. A. Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.* **56**, 391–412 (2008).
32. Lemey, P. et al. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2. *PLoS Pathog.* **10**, e1003932 (2014).
33. Suchard, M. A. & Rambaut, A. Many-core algorithms for statistical phylogenetics. *Bioinformatics* **25**, 1370–1376 (2009).
34. Gräf, T. et al. Contribution of epidemiological predictors in unraveling the phylogeographic history of HIV-1 subtype C in Brazil. *J. Virol.* **89**, 12341–12348 (2015).



Extended Data Figure 1 | See next page for caption.

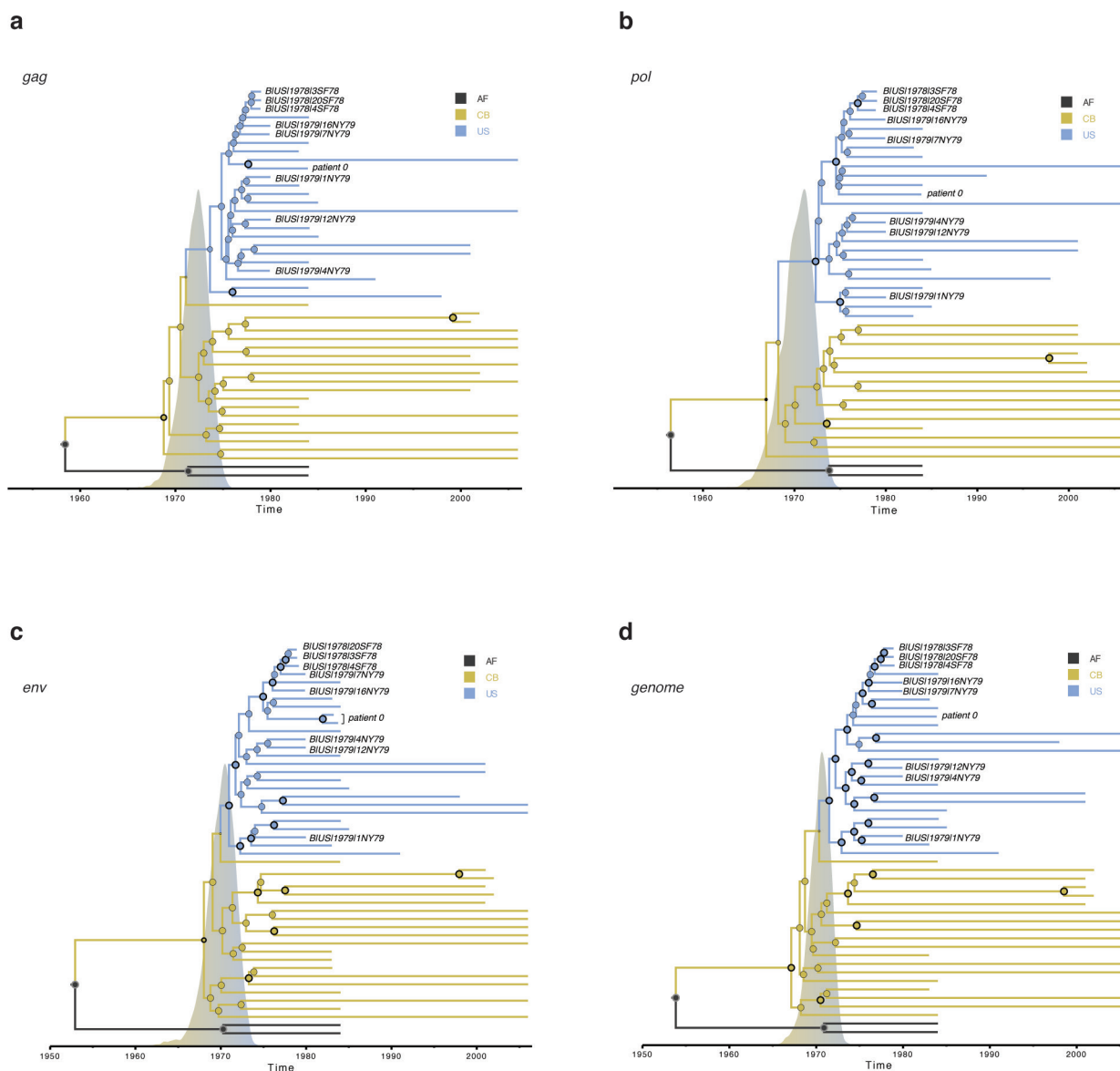
Extended Data Figure 1 | Jackhammering schematic and primer panels and pools. a–d, Detection and amplification of target RNA molecules in old, degraded, low-titre samples. For the purposes of illustration, consider a tube with 10^{13} RNA molecules, but (because of the low RNA quality) only one molecule that is (i) capable of being primed by the given reverse primer(s) and (ii) long enough to form a 200-bp product. **a,** Conventional RT–PCR with a long amplification product, oversized for a sample with RNA less than ~ 200 bases in length. **b,** RT–PCR with a shorter amplification product. **c,** Use of multiple primer pairs to increase the chance of at least one PCR-positive result. **d,** The jackhammering approach, which overcomes the problems encountered in **a–c** by (i) targeting an extensive panel of short amplicons appropriately sized

to the level of RNA survival in the sample, (ii) conducting reverse transcription with pools of primer pairs that amplify discrete, non-overlapping genomic regions, and (iii) employing a multiplex pre-amplification step, in the tube with the reverse transcription product, to generate sufficient DNA to ensure that each aliquot from it contains numerous template molecules for final PCR amplification. In this schematic, we show just two primer pairs per pool, but we used pools of ten pairs with our largest primer panels (shown in **e**, HXB2 numbering along HIV-1 genome). With a 10 primer-pair pool, and 10 final reactions, one can reliably recover 10 bands for sequencing. Five such pools (one entire panel of 50 pairs), allows complete HIV-1 genome recovery even in heavily degraded samples.

a**b**

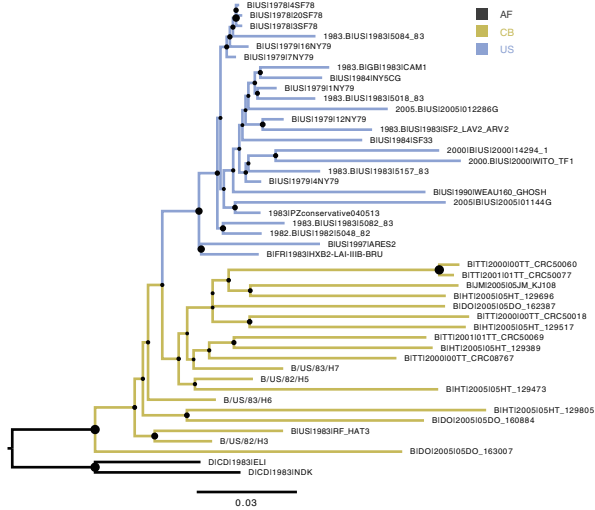
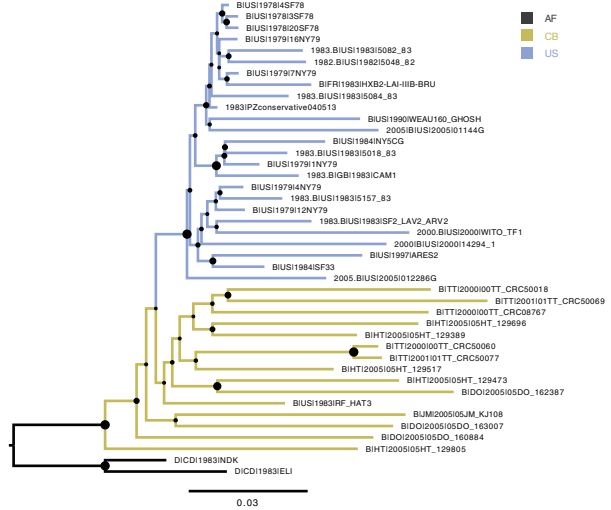
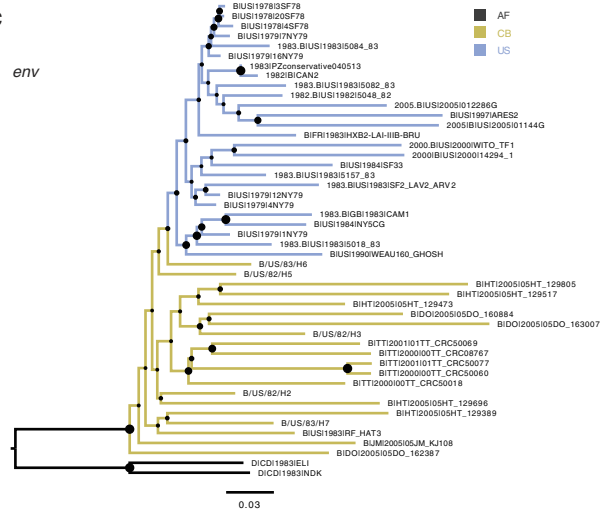
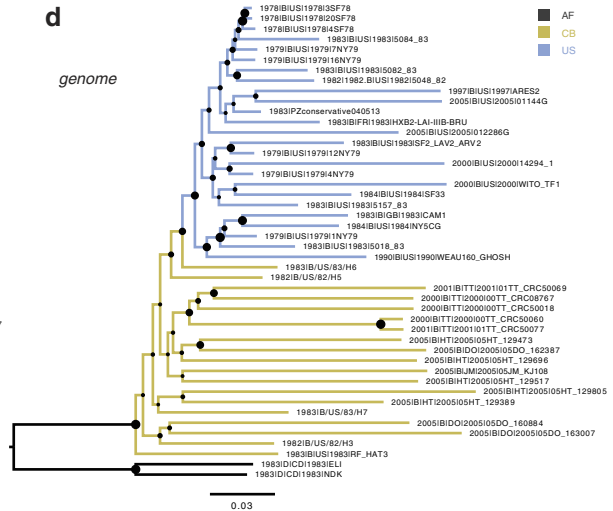
Extended Data Figure 2 | Maximum clade credibility (MCC) tree summaries of Bayesian spatio-temporal reconstructions based on complete HIV-1 genome data. a, 'full genome 46', b, 'full genome 38'. The tips of the trees correspond to the year of sampling while the branch (and node) colours reflect location: the sampling location for the tip branches and the inferred location for the internal branches. AF, Africa;

CB, Caribbean; US, the United States; CA, California, GA, Georgia; NY, New York. The diameters of the internal node circles reflect posterior location probability values. Thick outer circles represent internal nodes with posterior probability values > 0.95 . Gray bars indicate the 95% credibility intervals for the internal node ages. The tree in **b** represents the fully annotated version of the tree in Fig. 1 in the main text.



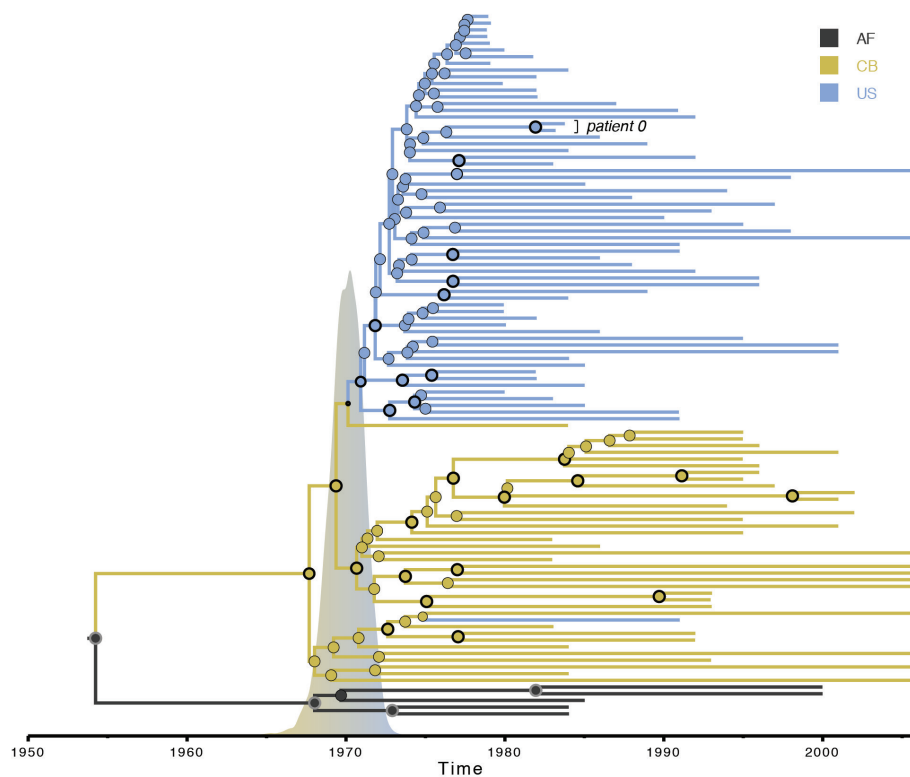
Extended Data Figure 3 | Maximum clade credibility (MCC) tree summaries of Bayesian spatio-temporal reconstructions based on different genome region data sets. MCC trees for the same strains are shown for **a**, *gag*, **b**, *pol*, **c**, *env* and **d**, the complete genome. The tips of the trees correspond to the year of sampling while the branch (and node) colours reflect location: the sampling location for the tip branches and the inferred location for the internal branches. AF, Africa; CB, Caribbean;

US, the United States. Tip labels are provided for the newly obtained archival HIV-1 genomes. The diameters of the internal node circles reflect posterior location probability values. Thick outer circles represent internal nodes with posterior probability support >0.95. We also depict the posterior probability densities for the time of the introduction event from the Caribbean into the US on the time scale of the trees.

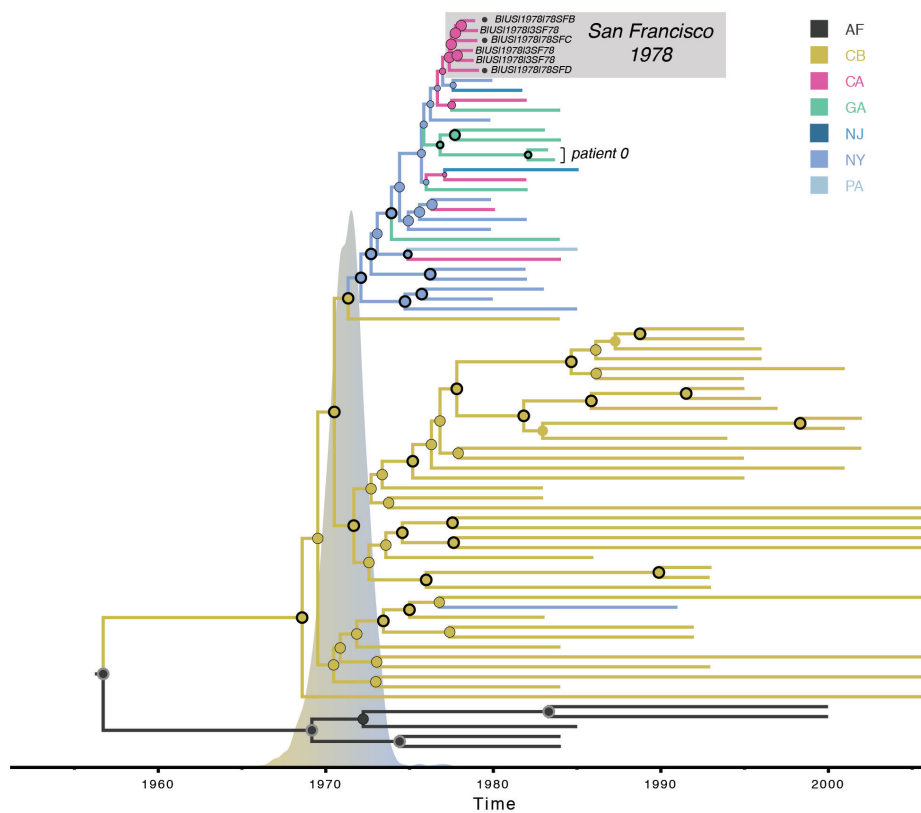
a*gag***b***pol***c***env***d***genome*

Extended Data Figure 4 | Maximum likelihood phylogenies for the different genome region data sets. a, *gag*, b, *pol*, c, *env* and d, the complete genome. We analysed the same data sets as in Extended Data Fig. 3. The diameters of the internal node circles reflect bootstrap support values. We manually coloured the branches in a similar way as for the Bayesian phylogeographic reconstructions.

a



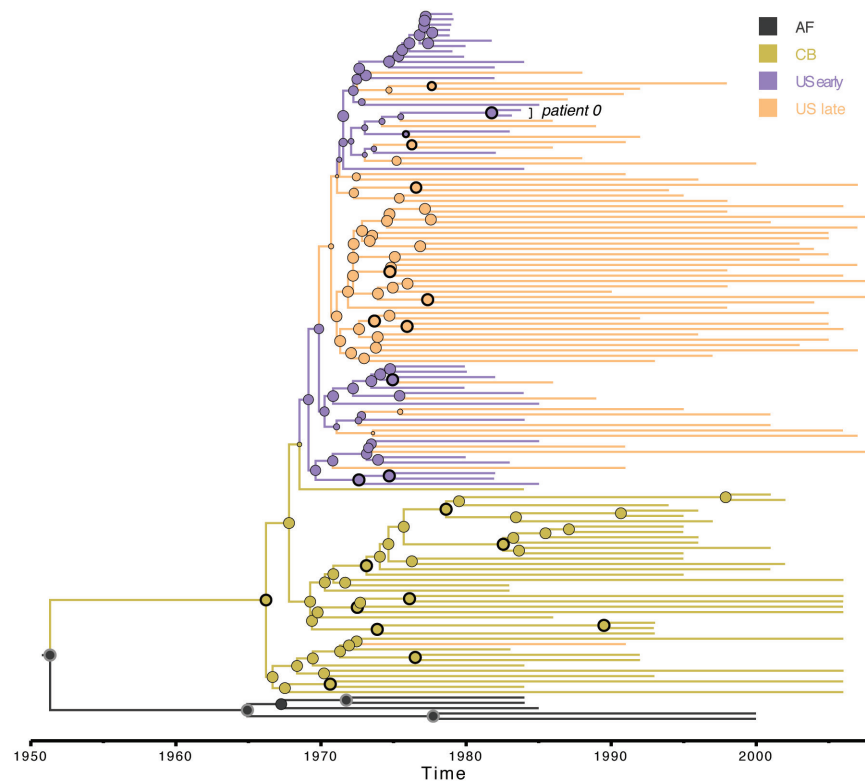
b



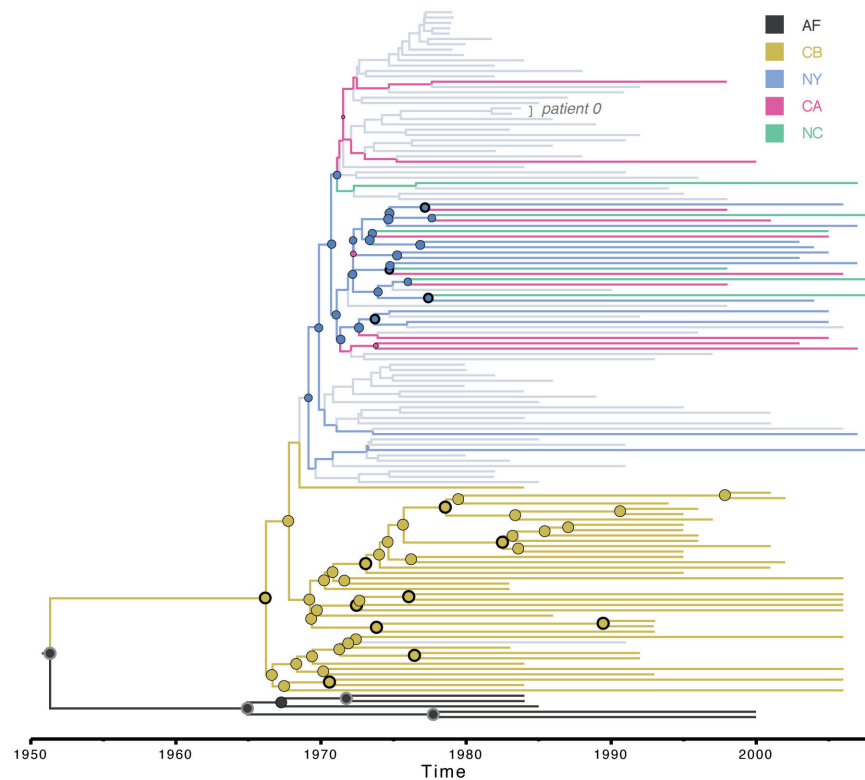
Extended Data Figure 5 | Maximum clade credibility (MCC) tree summaries of Bayesian spatio-temporal reconstructions based on different *env* data sets. a, 'env 105'; b, 'env 74'. The tips of the trees correspond to the year of sampling while the branch (and node) colours reflect location: the sampling location for the tip branches and the inferred location for the internal branches. AF, Africa; CB, Caribbean; US, the United States, CA, California; GA, Georgia; NJ, New Jersey,

NY, New York; PA, Pennsylvania. The diameters of the internal node circles reflect posterior location probability values. Thick outer circles represent internal nodes with posterior probability support >0.95. We also depict the posterior probability density for the time of the introduction event from the Caribbean into the U.S. on the time scales of the trees. The three partial *env* sequences from SF in 1978 (ref. 10) are highlighted with bullets.

a

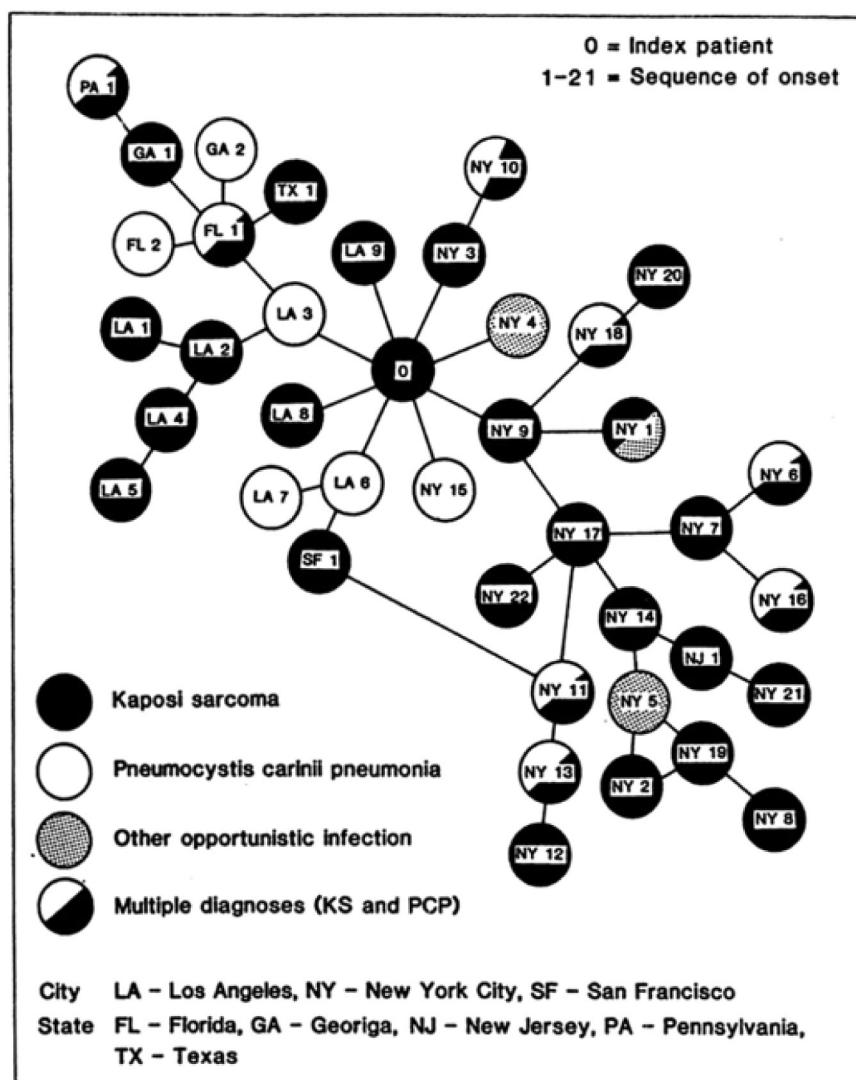


b

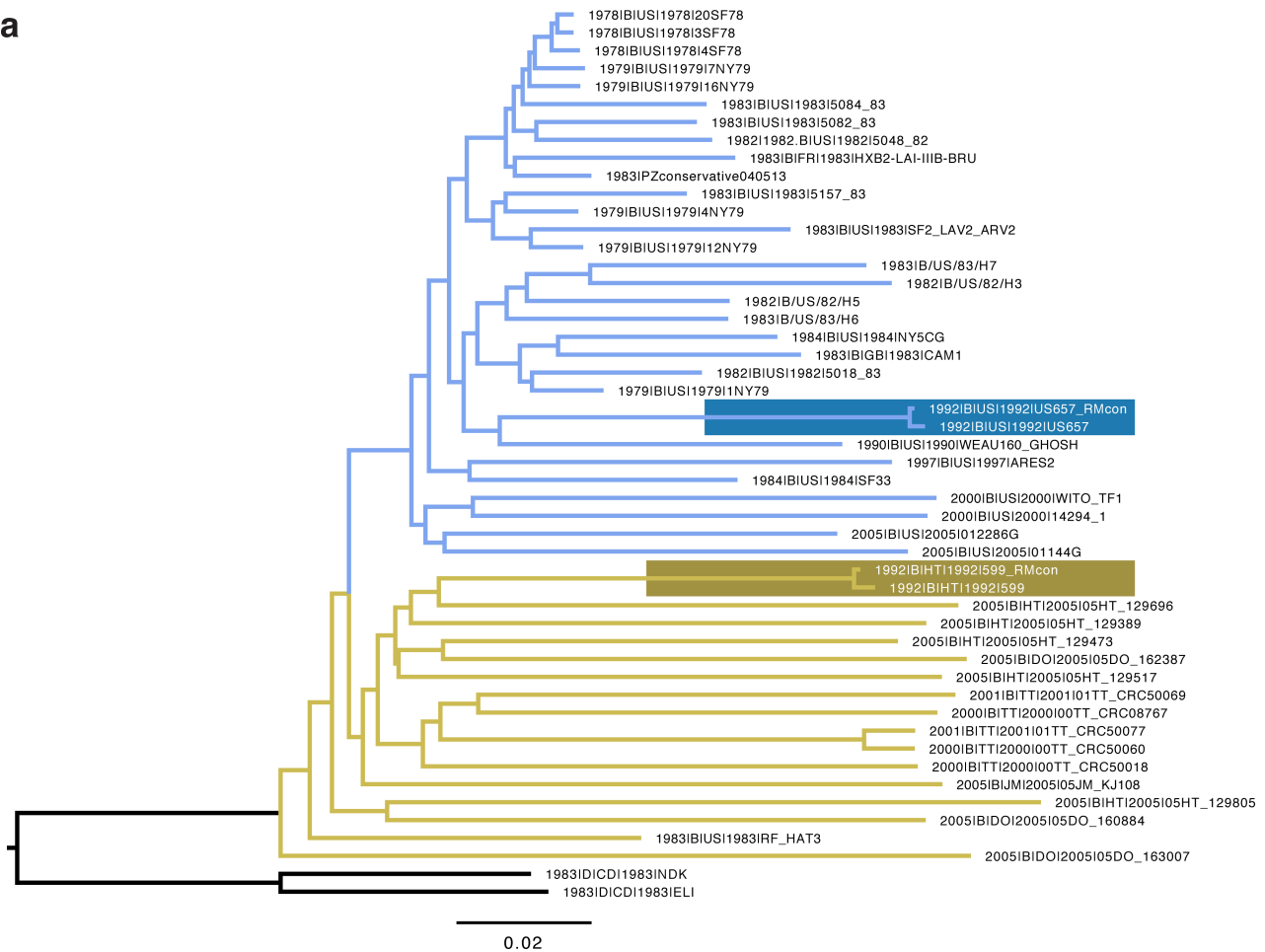
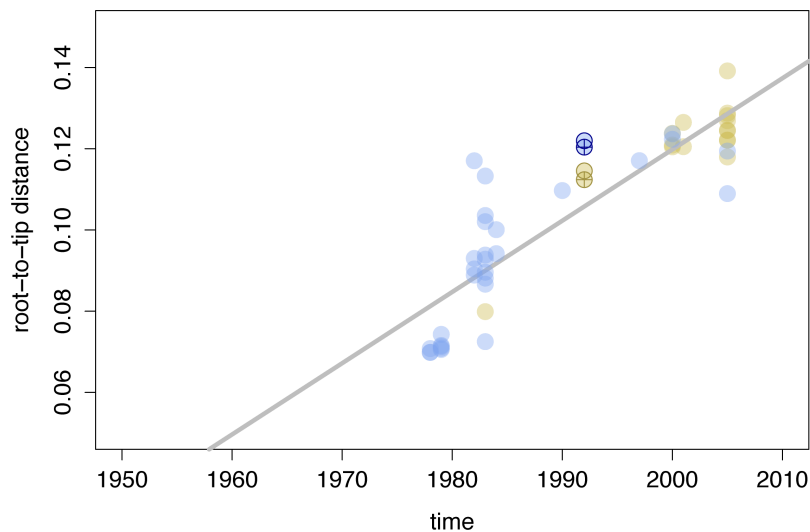


Extended Data Figure 6 | Maximum clade credibility (MCC) tree summaries of Bayesian spatio-temporal reconstruction comparing early and late strains. a, 'env 133'; b, only 'late' sequences from 'env 133'. In **a**, we classified US sequences as 'early' or 'late' depending on whether they were sampled before or after (and including) 1985. In **b**, the analysis was conducted on an empirical tree distribution of 'env 133' from which we pruned early US sequences (in grey), but we still annotate the reconstruction on the complete phylogenies for reference. The tips of

the tree correspond to the year of sampling while the branch (and node) colours reflect location: the sampling location for the tip branches and the inferred location for the internal branches. AF, Africa; CB, Caribbean; US early, the United States sampled <1985; US late, the United States sampled in or after 1985; CA, California; GA, Georgia; NC, North Carolina, NY, New York. The diameters of the internal node circles reflect posterior location probability values. Thick outer circles represent internal nodes with posterior probability support >0.95.

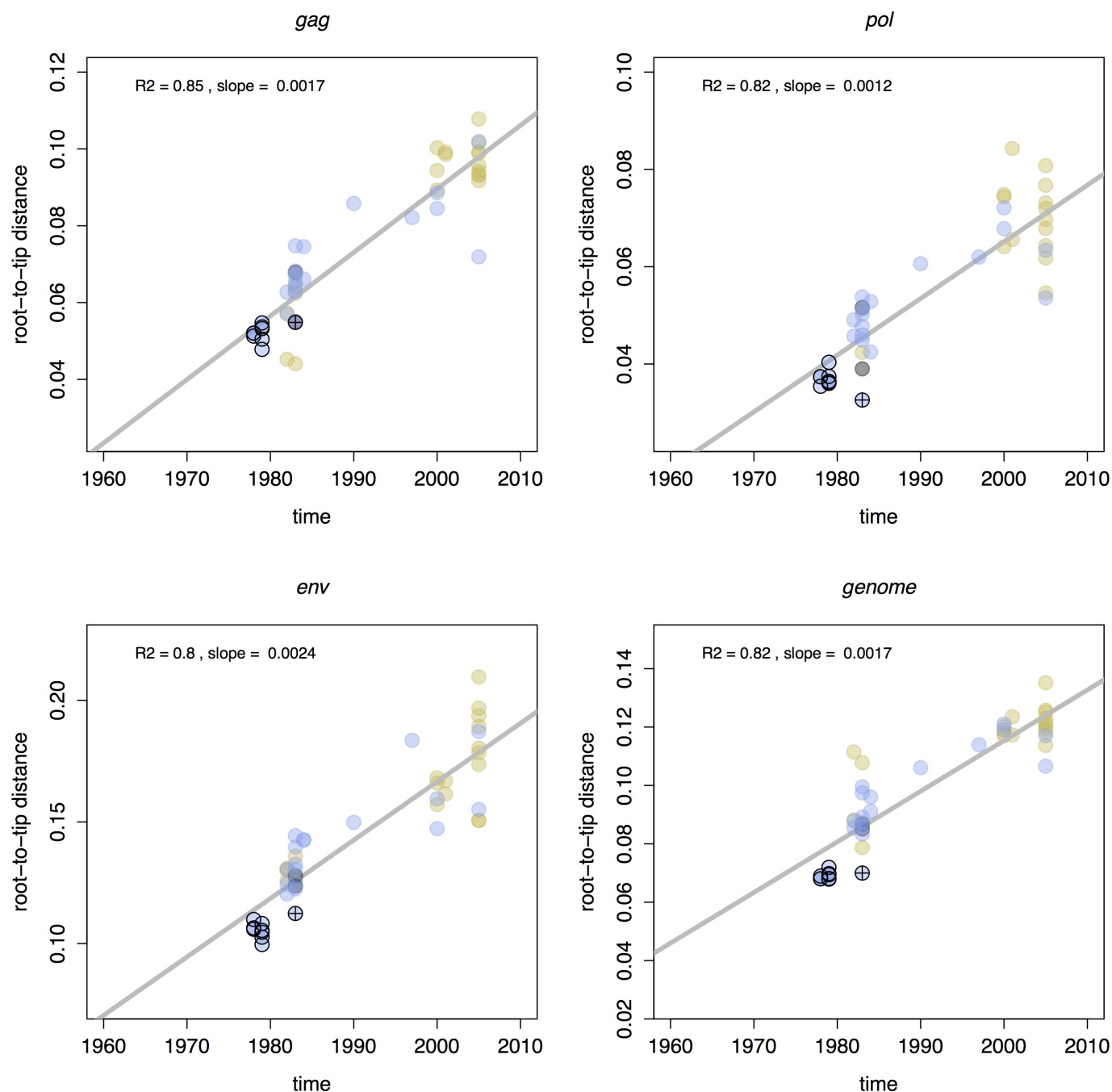


Extended Data Figure 7 | A cluster of 40 early AIDS patients linked through sexual contact. Reprinted from figure 1 of ref. 5 with permission from Elsevier.

a**b**

Extended Data Figure 8 | Jackhammering validation with reference viruses. **a**, The consensus sequences for primer panels HIVM and HIVR ('RMcon' suffix) were included, with previously published sequences for an US (US657) virus and a Haitian (HT599) virus, in a maximum likelihood tree. The two clusters of paired sequences are highlighted by coloured boxes. **b**, Plot of the root to tip genetic distance against sampling time for

the tree in **a**. The colours for the data points are consistent with those used for sampling locations in the phylogenies (the two African outgroup tips are not shown for clarity). The data points with black circles represent the published sequences while the data points with a target symbol represent the newly obtained sequences.



Extended Data Figure 9 | Plots of the root-to-tip genetic distance against sampling time for different genome region data sets (*gag*, *pol*, *env* and the complete genome). We used TempEst²⁷ to obtain exploratory regressions based on the maximum likelihood trees (Extended Data Fig. 4). Each data point represents a tip; colours are consistent with those

used for sampling locations in the phylogenies. The US data points with black circles represent the new genomes dating back to 1978–1979. The data point with the target symbol represents the Patient 0 genome. In each plot, we provide the R^2 for the regression and the slope, reflecting the evolutionary rate (in substitutions per site per year).

Extended Data Table 1 | Molecular clock, phylogeographic and recombination estimates for the different data sets

Data set	TMRCA (subtype B & D)	TMRCA (subtype B)	Location probability (subtype B)	Jump time (CB to US)	TMRCA (US subtype B)	Location probability (US subtype B)	Evolutionary rate	Rate, coefficient of variation	Phi test p-value
"full genome 46", ED Fig. 2 & ED Fig. 3	1953 (1946,1961)	1967 (1963,1970)	CB: > 0.99	1970 (1968,1973)	1972 (1969,1973)	US: > 0.99	0.0027 (0.0024,0.0030)	0.25 (0.20,0.31)	0.99
"full genome 38", Fig. 1 & ED Fig. 2	1955 (1946,1962)	1967 (1963,1970)	CB: 0.99	1971 (1968,1973)	1972 (1970,1974)	NY: > 0.99	0.0024 (0.0021,0.0027)	0.26 (0.20,0.32)	0.99
"gag", ED Fig. 3	1958 (1950,1964)	1969 (1964,1972)	CB: > 0.99	1972 (1969,1974)	1974 (1971,1975)	US: > 0.99	0.0023 (0.0020,0.0026)	0.23 (0.14,0.33)	0.77
"pol", ED Fig. 3	1956 (1947,1965)	1967 (1961,1972)	CB: 0.92	1970 (1966,1973)	1973 (1969,1974)	US: > 0.99	0.0015 (0.0013,0.0017)	0.29 (0.20,0.37)	0.21
"env", ED Fig. 3	1953 (1943,1962)	1968 (1964,1972)	CB: > 0.99	1970 (1966,1973)	1971 (1968,1974)	US: 0.99	0.0037 (0.0032,0.0043)	0.25 (0.16,0.34)	< 0.01
"env, recomb. free"	1952 (1940,1961)	1968 (1964,1972)	CB: 0.99	1970 (1966,1973)	1971 (1967,1973)	US: 0.99	0.0039 (0.0031,0.0047)	0.26 (0.18,0.35)	0.59
"env 105", ED Fig. 5	1954 (1947,1961)	1968 (1964,1971)	CB: > 0.99	1970 (1968,1972)	1971 (1969,1973)	US: > 0.99	0.0047 (0.0042,0.0052)	0.23 (0.18,0.28)	0.01
"env 105, recomb. free"	1955 (1947,1961)	1968 (1974,1970)	CB: > 0.99	1970 (1968,1972)	1971 (1969,1972)	US: > 0.99	0.0047 (0.0041,0.0053)	0.23 (0.18,0.28)	0.26
"env 74", ED Fig. 5	1957 (1948,1963)	1969 (1963,1971)	CB: > 0.99	1971 (1969,1973)	1972 (1969,1974)	NY: 0.97	0.0044 (0.0038,0.0050)	0.28 (0.21,0.36)	< 0.01
"env 74, recomb. free"	1957 (1948,1964)	1969 (1964,1972)	CB: 0.99	1971 (1968,1973)	1972 (1970,1974)	NY: 0.97	0.0046 (0.0038,0.0054)	0.31 (0.23,0.39)	0.91
"env 133", ED Fig. 6†	1952 (1944,1958)	1966 (1963,1969)	CB: 0.99	1969 (1966,1971)	1969 (1967,1971)	NY: 0.67	0.0045 (0.0041,0.0048)	0.20 (0.16,0.23)	0.76

*The recombination free ('recomb. free') data sets were obtained by deleting the minor recombinant regions from the putative recombinants identified using RDP4.

†The empirical trees from the 'env 133' analysis were used for two different ancestral reconstructions (Extended Data Fig. 6); here we list the location estimates for the analysis that considered different US states for the late samples (Extended Data Fig. 6b).

Single-cell RNA-seq identifies a PD-1^{hi} ILC progenitor and defines its development pathway

Yong Yu^{1*}, Jason C. H. Tsang^{1,2,3*}, Cui Wang^{1,4*}, Simon Clare¹, Juexuan Wang¹, Xi Chen¹, Cordelia Brandt¹, Leanne Kane¹, Lia S. Campos¹, Liming Lu⁵, Gabrielle T. Belz^{6,7}, Andrew N. J. McKenzie⁸, Sarah A. Teichmann^{1,9}, Gordon Dougan^{1,10} & Pentao Liu¹

Innate lymphoid cells (ILCs) functionally resemble T lymphocytes in cytotoxicity and cytokine production but lack antigen-specific receptors, and they are important regulators of immune responses and tissue homeostasis^{1,2}. ILCs are generated from common lymphoid progenitors, which are subsequently committed to innate lymphoid lineages in the α -lymphoid progenitor, early innate lymphoid progenitor, common helper innate lymphoid progenitor and innate lymphoid cell progenitor compartments^{3–8}. ILCs consist of conventional natural killer cells and helper-like cells (ILC1, ILC2 and ILC3)⁹. Despite recent advances^{1,2,10}, the cellular heterogeneity, developmental trajectory and signalling dependence of ILC progenitors are not fully understood. Here, using single-cell RNA-sequencing (scRNA-seq) of mouse bone marrow progenitors, we reveal ILC precursor subsets, delineate distinct ILC development stages and pathways, and report that high expression of programmed death 1 (PD-1^{hi}) marked a committed ILC progenitor that was essentially identical to an innate lymphoid cell progenitor. Our data defined PD-1^{hi}IL-25R^{hi} as an early checkpoint in ILC2 development, which was abolished by deficiency in the zinc-finger protein *Bcl11b* but restored by IL-25R overexpression. Similar to T lymphocytes, PD-1 was upregulated on activated ILCs. Administration of a PD-1 antibody depleted PD-1^{hi} ILCs and reduced cytokine levels in an influenza infection model in mice, and blocked papain-induced acute lung inflammation. These results provide a perspective for exploring PD-1 and its ligand (PD-L1) in immunotherapy, and allow effective manipulation of the immune system for disease prevention and therapy.

We purified individual Lin[−]Flt3^{lo/−}IL-7R α ^{lo/+} α 4 β 7⁺ bone marrow cells^{5,6,10} for scRNA-seq analysis (4 × 96 wild type and 2 × 96 *Bcl11b*-deficient) (Extended Data Fig. 1a). After sequencing quality control (Extended Data Fig. 1b–j), 325 wild-type cells (172 CD244⁺CD25[−], 84 CD244[−]CD25[−] and 69 CD244[−]CD25⁺), and 172 *Bcl11b*-deficient cells (Lin[−]Flt3[−]IL-7R α ⁺ α 4 β 7⁺) were further analysed. Comparison of External RNA Controls Consortium (ERCC) RNA spike-ins in libraries did not reveal significant differences between experiments, indicating minimal batch bias (Extended Data Fig. 1g).

In total, 8,758 genes showed significant biological expression variability¹¹ (Extended Data Fig. 1h) and were used for dimensional reduction analysis by *t*-distributed stochastic neighbour embedding (t-SNE). Distinct aggregation pattern of ILCs was observed in the wild-type dataset (Fig. 1a), where cells were roughly grouped into 10 clusters, probably representing bone marrow cells of distinct lineages and/or developmental stages. Two major distinct subpopulations were immediately evident on the basis of expression of ILC regulators *Id2*,

Gata3 and *Il7r*: Id2^{lo}Gata3^{lo}IL-7R α ^{lo} (clusters 1, 2 and 3, denoted as C1–C3) and Id2^{hi}Gata3^{hi}IL-7R α ^{hi} (C6–C10) (Fig. 1b–d). Expressions of megakaryo-erythrocytic and myeloid regulators in the Id2^{lo}Gata3^{lo}IL-7R α ^{lo} (C1–C3) suggested that they are non-ILCs (Fig. 1d, Extended Data Figs 2, 3). *Bcl11a* is expressed in common lymphoid progenitors, but not in T cells or conventional NK cells (cNKs)¹². *Bcl11a* decreased in C4 and was very low in other clusters (Fig. 1b).

Many C4 and C5 cells expressed *Il7r* or *Gata3* (Fig. 1b–d). C4 had high *Notch1* and detectable *Flt3*, and a small number of C4 cells expressed *Nfil3* (Fig. 1d), which is important in α lymphoid progenitors (αLPs) development^{7,13,14}. C4 cells expressed no *Tcf7* (that encodes TCF-1), *Rorc* (ROR γ t), *Cxcr5*, *Id2* or *Zbtb16* (PLZF) (Fig. 1c, d), and therefore are probably αLP1 progenitors^{3,7}. By contrast, C5 had high *Tcf7*, *Gata3*, *Tox*, *Tox2* (ref. 15) and *Ets1* (ref. 16) expression, but did not express *Bcl11a* or *Flt3* (Fig. 1b–d). C5 cells were thus reminiscent of early innate lymphoid progenitors (Lin[−]TCF-1⁺Thy-1[−]IL-7R α ^{lo/+})⁶, which have potential to all ILC lineages. Notch signalling induces TCF-1 expression in T-cell and ILC development^{6,17}. Concordantly, *Notch1* peaked in C4 before *Tcf7* and *Hes1* induction in C5 (Extended Data Fig. 2). We only observed a small number of C5 cells in the wild-type dataset, possibly owing to our preferential sorting for IL-7R α ^{lo/+} cells.

Common helper innate lymphoid progenitors (CHILPs) are defined as Lin[−]Id2⁺IL-7R α ⁺ α 4 β 7⁺CD25[−] (ref. 5). Almost all C6–C10 cells highly expressed *Id2* (Fig. 1b–d). C6–C9 cells expressed *Il7r*, *Itga4* and *Itgb7* (encoding integrin α 4 β 7), but not *Il2ra* (CD25), and probably represented the CHILP compartment. Specifically, C6 cells expressed high *Zbtb16* and *Il7r*, *Kit*, *Itga4* and *Itgb7*, but low *Myc* (Fig. 1c, Extended Data Fig. 2), which is consistent with innate lymphoid cell progenitor (ILCP) phenotype (Lin[−]IL-7R α ⁺c-Kit⁺ α 4 β 7^{hi}PLZF^{hi}) (ref. 4) and the finding that *Myc* is suppressed by PLZF¹⁸.

C7a cells highly expressed many ILC2 genes such as *Bcl11b*, *Icos*, *Rora* and *Gata3*, and notably, C8–C10 cells gradually increased in expression of ILC2 genes including *Il2ra*, *Il1r1* (IL-33R), *Bmp7* and *Pparg* (Fig. 1d, Extended Data Fig. 2). Therefore, ILC2 development probably follows the trajectory from C7a to C10. Through scRNA-seq, we discovered potential new ILC2 genes that were expressed primarily in C8–C10 cells: for example, *Itgb3*, *Pbxip1*, *1700113H08Rik* (in C8 cells); *Ccr8*, *Gcl* (in C9 cells); and *Cish* (in C10 cells) (Extended Data Fig. 3). In contrast to C7a cells, C7b cells expressed genes enriched in ILC1s and/or cNKs, for example, *Tbx21* (T-bet), *Il2rb* (CD122), *Ncr1* (NKp46), *Cxcr3* and *Ctsw* (Extended Data Figs 2, 3). ROR γ t⁺ ILC3 precursors are rare in the adult mouse bone marrow^{4,5}, whereas ILC2 precursors are well-represented¹⁹. A small number of C7c cells exhibited features of ILC3 cells: high expression of *Rorc*, *Ahr*, *Ccr7* and *Cxcr5* (Extended Data Figs 2, 3).

¹Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, UK. ²Department of Chemical Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, New Territories, Hong Kong SAR, China. ³Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong SAR, China. ⁴Institute of Animal Husbandry and Veterinary Science, Shanghai Academy of Agricultural Sciences, Shanghai 201106, China. ⁵Shanghai Institute of Immunology, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China. ⁶The Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria 3052, Australia. ⁷Department of Medical Biology, University of Melbourne, Melbourne, Victoria 3010, Australia. ⁸Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. ⁹European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ¹⁰Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0SP, UK.

*These authors contributed equally to this work

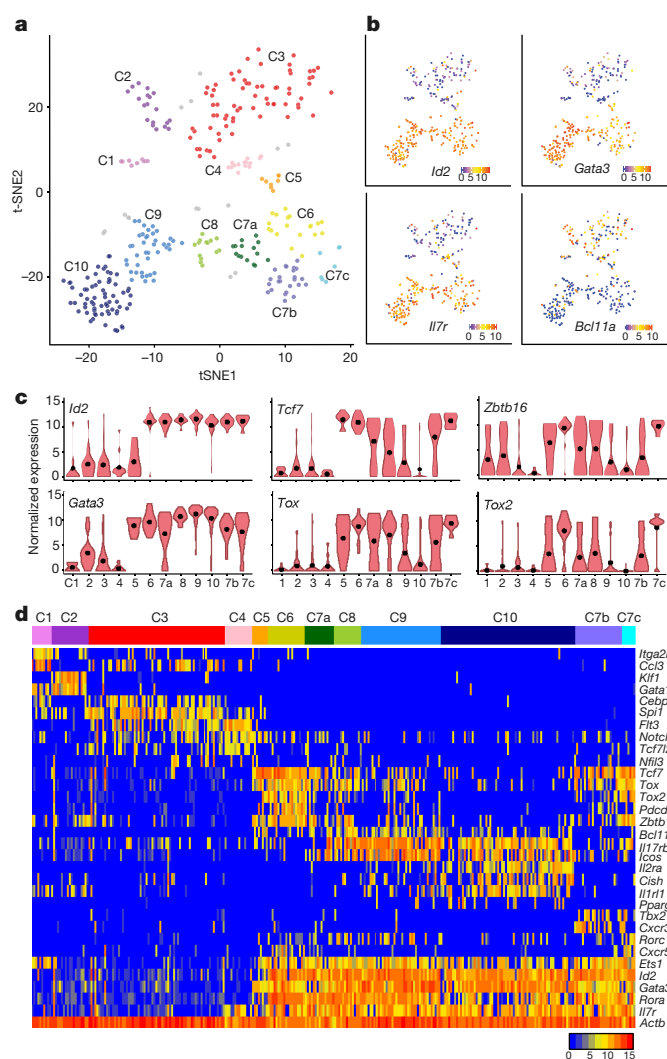


Figure 1 | scRNA-seq analysis of innate lymphoid progenitors. **a**, Biaxial t-SNE clustering plot of 325 wild-type Lin⁺Flt3^{lo}/IL-7R α ^{+/lo} α β γ ⁺ bone marrow cells. Each point represents an individual cell. Cells were grouped into 10 distinct clusters (C1–C10, colours indicated). **b**, Distribution of gene expression in t-SNE plot. The colour key indicates the expression level. **c**, Violin plots comparing expression of key ILC regulators in different clusters. The y axis indicates the log₂ (normalized count + 1) expression levels. The black point indicates the mean of expression level. **d**, Heat map of the expression levels of selected genes in different clusters.

Notably, we observed distinct T-cell-receptor transcript expression demarcation between C1–C4 (non-ILCs and uncommitted progenitors) and C5–C10 (ILC progenitors) (Extended Data Fig. 4).

ILC progenitor isolation currently necessitates the use of multiple cell surface markers and genetic reporters (*Id2*, *Tcf7* or *Zbtb16*)^{4–6}. We sought to find novel ILC markers that can remove this obstacle. Closer examination of C6 revealed high levels of *Pdcd1* (Fig. 2a), which encodes PD-1, a cell surface receptor and a member of the immunoglobulin super family. Notably, *Pdcd1* and *Zbtb16* were positively correlated in C6 cells (Extended Data Fig. 5a). *Pdcd1*-expressing C6 cells also expressed *Id2*, *Tcf7*, *Tox* and *Gata3*, but lacked *Flt3* or markers of mature ILCs (Extended Data Fig. 5b). Unlike activated T cells, C6 cells did not express other inhibitory or activation molecules²⁰, or PD-1 ligand genes *Pdcd1lg1* (also known as *Cd274*) or *Pdcd1lg2* (Extended Data Fig. 5b), suggesting that high *Pdcd1* expression was unique to ILC progenitors in the Lin⁺ compartment.

We then isolated Lin⁺PD-1^{hi} cells (PD-1^{hi} hereafter) and confirmed strong expression of multi-lineage ILC regulators: PLZF, TCF-1, TOX

and GATA3 (refs 4, 6, 15, 21) (Fig. 2b). They were positive for ILC progenitor markers but not Flt3 or lineage-committed markers (Extended Data Fig. 5c). PD-1^{hi} cells showed *in vivo* developmental potential in adoptive transfer experiments in lymphoid *Rag2*^{−/−}*Il2rg*^{−/−} mice. In addition, engrafted control common lymphoid progenitors produced B cells, T cells and all ILCs (cNK: NK1.1⁺CD49a[−]CD49b⁺, ILC1s, ILC2s and ILC3s, including lymphoid-tissue-inducer cells (LTi)), whereas PD-1^{hi} cells generated liver NK1.1⁺CD49a⁺CD49b[−] ILC1 cells, ILC2 and ILC3 cells, and a small number of cNKs, but no B, T or CCR6⁺ LTi cells (Extended Data Fig. 5d). Within the experimental time window, common lymphoid progenitors produced PD-1^{hi} cells in the bone marrow, but all of the donor PD-1^{hi} cells lost PD-1 expression in the recipients, indicating their differentiation (Extended Data Fig. 5d).

To determine the precursor–descendant relationship, we seeded individual PD-1^{hi} bone marrow cells on stromal cells and assayed their progenies by flow cytometry. Individual PD-1^{hi} cells generated mixtures of ILC lineages (double (two) or triple (three) lineages of ILC1s, ILC2s and ILC3s) in 22 out of 98 wells (22.4%) (Extended Data Fig. 5e), demonstrating that they were enriched in both unipotent and multipotent ILC progenitors.

We next directly compared PD-1^{hi} cells to PLZF^{hi} cells (ILCP) or CHILPs. In the *Zbtb16*^{GFPcre} reporter mouse, in which *Zbtb16* (PLZF) expression is tracked by GFP⁴, the majority of PD-1^{hi} cells also expressed high PLZF as detected by flow cytometry (Fig. 2c). Co-transplantation confirmed their similar *in vivo* developmental potential (Extended Data Fig. 6a). Therefore, innate lymphoid cell progenitors could now be isolated by simple PD-1 staining, without relying on a transgenic reporter. In the *Id2*^{GFP} reporter mouse⁵, PD-1^{hi} cells were all CHILPs (Fig. 2d), whereas about 50% of CHILPs were PD-1^{hi} cells (Fig. 2e). Many PD-1^{lo} cells in CHILP increasingly expressed lineage markers and downregulated c-Kit (Fig. 2f), indicating potential lineage commitment.

To estimate the heterogeneity in ILC progenitor compartments, we crossed *Id2*^{GFP} mice with *Bcl11b*^{tdTomato} conditional knockout mice (Extended Data Fig. 6b). Flow cytometry analyses detected *Bcl11b*⁺ cells in about 50% of CHILPs and 40% of PD-1^{hi} cells, respectively (Extended data Fig. 6c, d). PD-1^{hi}*Bcl11b*⁺ bone marrow cells produced ILC2 only, whereas in clonal assays PD-1^{hi}*Bcl11b*[−] cells generated mixtures of ILCs of double and triple lineages in as many as 50% of wells and showed ILC1–3 potential *in vivo* (Extended Data Figs 5e, 6e).

To study the PD-1^{hi} compartment further, we performed scRNA-seq analysis of a further 184 PD-1^{hi} cells. Clustering analysis showed polarized expression of *Il17rb* (46 out of 184, 25%) (Extended Data Fig. 7a). The differential *Il17rb* expression and spatial proximity in t-SNE were largely consistent with that from hierarchical clustering (Extended Data Fig. 7b), revealing that *Pdcd1* expression was associated with *Zbtb16*, *Tcf7*, *Tox* and *Tox2*, and that those cells with low levels of *Pdcd1* or *Tox* expressed higher *Il17rb*, indicating that they were primed or committed to an ILC lineage.

In the initial scRNA-seq dataset, prominent expression of *Il17rb* was first detected in C7a and reached the highest levels in C8 and C9 (Fig. 3a). Gene set enrichment analysis revealed that from C7a to C8, ILC2 signature was increased, whereas ILC1 and cNK signature was downregulated (Fig. 3b), and that *Il17rb* and *Bcl11b* were among those showing spike or peak expression at C7a to C8 (Extended Data Fig. 8a).

Experimentally, about 20–30% of the PD-1^{hi} cells were IL-25R⁺, of which all were *Bcl11b*⁺, and clonal differentiation and adoptive transfer experiments confirmed that IL-25R⁺ cells were ILC2-restricted (Extended Data Fig. 8b–d). These data revealed that IL-25R marked early ILC2 progenitors in PD-1^{hi} cells, and that *Bcl11b* is essential for ILC2 development^{22,23}. To determine the exact stage that ILC2 development becomes defective in the absence of *Bcl11b* and investigate the relationship between *Bcl11b* and *Il17rb*, we generated the *Vav*-Cre-*Bcl11b*^{fl/fl} mouse where *Bcl11b* was specifically deleted in hematopoietic cells and found that neither PD-1^{hi}IL-25R⁺ (Fig. 3c, d) nor peripheral ILC2s (Extended Data Fig. 8e) were present. scRNA-seq

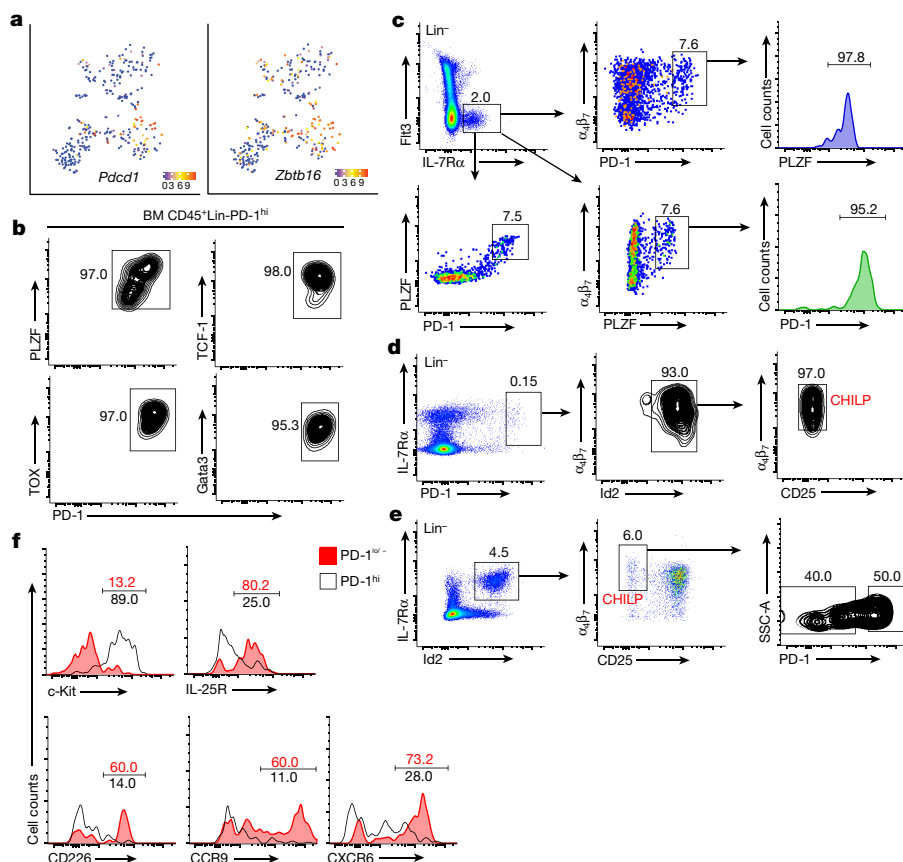


Figure 2 | PD-1 high expression marks innate lymphoid progenitors. **a**, Distribution of *Pcd1* and *Zbtb16* expression in t-SNE plots. **b**, Expression of key ILC regulators in PD-1^{hi} cells. Protein expression was measured by intracellular antibody staining. **c**, Comparison of PD-1^{hi} and PLZF^{hi} cells in the *Zbtb16*^{GFPcre} reporter mice ($n = 4$). **d**, FACS plots show all PD-1^{hi} cells were CHILPs using the *Id2*^{GFP} mice ($n = 6$). **e**, PD-1^{hi} cells are a subset of the CHILP compartment ($n = 6$). **f**, Detection of ILC lineage markers on PD-1^{lo/-} or PD-1^{hi} cells in CHILPs. Data are representative of two (**b**, **c**) or three (**d**–**f**) independent experiments.

of *Bcl11b*-deficient bone marrow cells (Lin⁻Flt3⁻IL-7Rα⁺αβ7⁺) further revealed the lack of ILC2 progenitors corresponding to those in C8–C10 of the wild-type bone marrow (Fig. 3e). Intriguingly, the

mutant C6 cells still expressed high levels of key ILC regulators *Gata3*, *Tcf7*, *Tox* and *Id2* (Extended Data Fig. 9a, b), and retained ILC2 signature similar to wild-type cells (Fig. 3f). Therefore, *Bcl11b*-deficiency

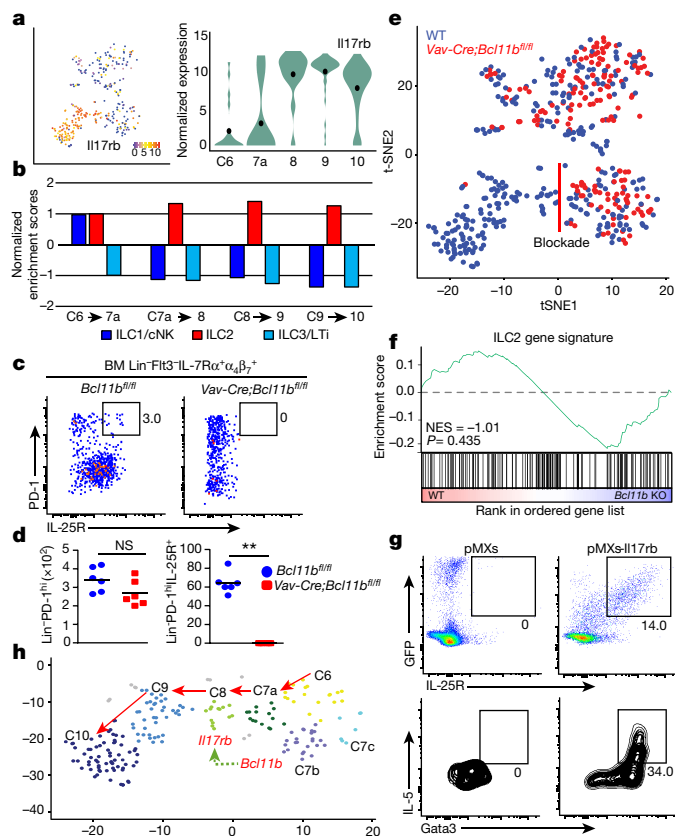


Figure 3 | scRNA-seq dissection of early ILC2 development.

a, Distribution of *Il17rb* in t-SNE biaxial plot and violin plot in C6–C10 clusters. **b**, Gene set enrichment analysis of expression changes. Stepwise comparisons between clusters of each transition step were performed and tested for enrichment of ILC1/cNK, ILC2 and ILC3/LTi signatures. The y axis indicates the normalized enrichment score, the x axis indicates the transition steps tested. Positive and negative enrichments indicate a relative enrichment or depletion in ILC signature genes at that transition, respectively. **c**, FACS analysis of PD-1^{hi}IL-25R⁺ cells in the control or *Vav-Cre-Bcl11b*^{fl/fl} mice ($n = 3$ per genotype). **d**, PD-1^{hi} and PD-1^{hi}IL-25R⁺ cell numbers in one femur of a control or mutant mouse ($n = 6$). $**P < 0.01$ (two-tailed *t*-test). **e**, t-SNE plot of sequenced wild-type and *Vav-Cre-Bcl11b*^{fl/fl} bone marrow cells. C8–C10 cells were missing in the mutant, indicating a developmental blockade. **f**, Gene set enrichment analysis comparing the ILC2 signature of C6 cells of wild-type and mutant bone marrow cells. **g**, Overexpression of *Il17rb* rescued ILC2 development of *Bcl11b*-deficient PD-1^{hi} cells *in vitro*. The infected cells were cultured on OP9-DL1 stromal cells and analysed two weeks later. **h**, The proposed ILC2 lineage development pathway. The red arrows indicate the developmental transition from one cluster to another. Data are representatives of three (**c**, **d**) or two (**g**) independent experiments.

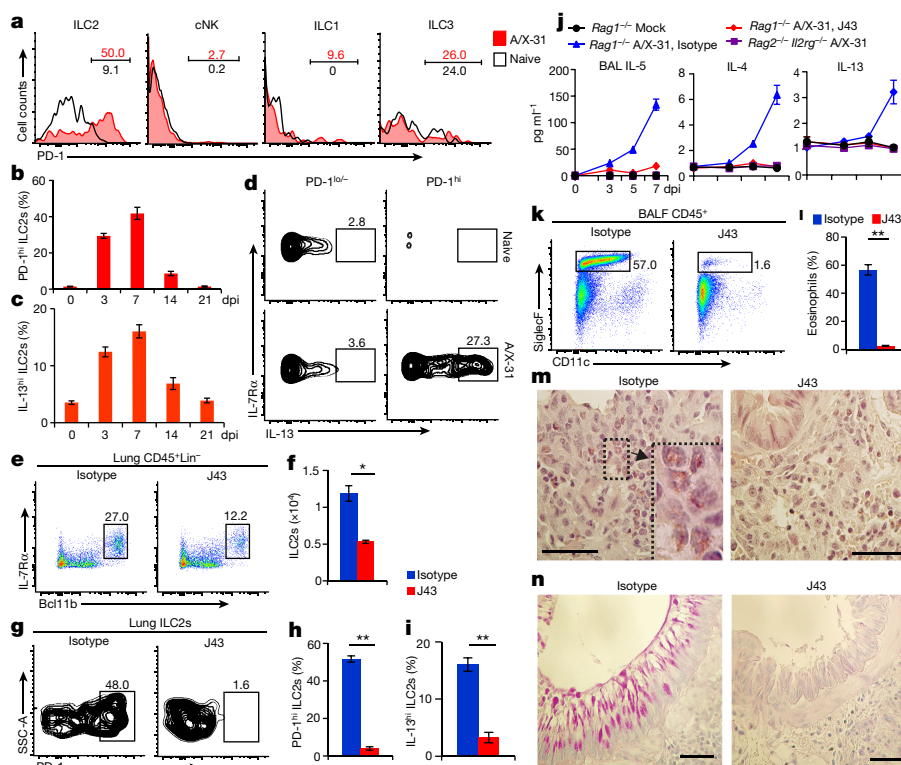


Figure 4 | PD-1^{hi} marks effector ILCs. **a**, PD-1^{hi} cells in the influenza A/X-31-infected (day 5) *Rag1*^{-/-} mice ($n = 3$). **b**, **c**, PD-1^{hi} ILC2s and IL-13^{hi} ILC2s in infected mice ($n = 3$ per time point). **d**, Only PD-1^{hi} cells in infected mice produced high levels of IL-13. IL-13 was detected in the *Il13*^{+/-tdTomato} mice ($n = 6$). **e–j**, Depletion PD-1^{hi} ILCs by J43 antibody in infected mice. An isotype control antibody was used as a negative control. Lung tissue was collected at the indicated time points ($n = 5$ per treated group). **e**, **f**, Lung ILC2s were quantified as shown (7 days after infection). ILC2s: Lin⁻CD45⁺IL-7R α ⁺Bcl11b⁺. **g**, **h**, Percentage of lung PD-1^{hi} ILC2s as shown (7 days after infection). **i**, Percentage of lung IL-13^{hi} ILC2s (7 days after infection). **j**, Type-2-related cytokines in BALF of the mice of different genotypes or treatments ($n = 3$ per group per time point): *Rag1*^{-/-} mice with mock infection; *Rag1*^{-/-} mice infected with A/X-31 and treated with either an isotype control or J43. *Rag2*^{-/-} *Il2rg*^{-/-} mice

infected with A/X-31 as the ILC-deficient control. **k–n**, Papain challenge of *Rag1*^{-/-} mice. The mice were pretreated with J43 or the isotype control for 3 days and administered with papain (intranasally) for 5 consecutive days. The lung tissues and BALF were collected at day 6 for analysis ($n = 5$ per treatment). **k**, **l**, Eosinophils (SiglecF⁺CD11c⁻) in the BALF were analysed as shown. **m**, Haematoxylin and eosin staining of lung sections. Inflammation with groups of eosinophils was observed in the isotype control lung. J43 treatment shows a reduction in the number of eosinophils. **n**, Periodic-acid-Schiff staining of lung sections. Positively stained goblet cells were found in the isotype control lung but not in the J43-treated lung in an equivalent section of respiratory epithelium. Scale bar (**m**, **n**), 50 μ m. Error bars (**b**, **c**, **f**, **h**, **i**, **j**, **l**) denote s.e.m. Data are representatives of two (**a–n**) independent experiments. * $P < 0.05$, ** $P < 0.01$ (two-tailed t -test).

blocked ILC2 development at the C7a to C8 transition (Fig. 3e), where expression of *Bcl11b* and *Il17rb* peaked (Extended Data Fig. 8a). These observations led us to explore the possibility that overexpressing *Il17rb* may restore *Bcl11b*-deficient ILC2 development. We thus expressed transgenic *Il17rb* in *Bcl11b*-deficient PD-1^{hi} cells. The transgenic mutant cells had typical ILC2 markers (Extended Data Fig. 9c) and produced IL-5 (Fig. 3g). Therefore, in the ILC2 developmental trajectory (Fig. 3h), Bcl11b functions through IL-25R signalling at the early commitment checkpoint (C7a to C8). Upon IL-25 administration, the *Vav*-Cre-*Bcl11b*^{fl/fl} mice produced neither 'natural' ILC2 nor 'inflammatory' ILC2 cells²⁴ and did not show increased IL-5 levels (Extended Data Fig. 8f). On the other hand, NK1.1⁺NKp46⁺ ILCs were increased in the *Bcl11b*-deficient bone marrow and small intestinal lamina propria (siLP) (Extended Data Fig. 8e). In adoptive transfer experiments, the mutant PD-1^{hi} ILC progenitors developed into ILC1s and ILC3s, but not ILC2s (Extended Data Fig. 8g).

PD-1 is an inhibitory receptor on activated T cells and a checkpoint molecule. It is being exploited for re-activating cytotoxic T cells in cancer immunotherapy²⁵. We further expanded our scope to study PD-1 expression in tissue-resident ILCs. At steady-state, PD-1⁺ cells accounted for 20–40% ILC2s, 20–30% ILC3s and approximately 76% siLP LTi cells, but fewer cNK or ILC1 cells (Extended Data Fig. 10a, b). By contrast, influenza infection substantially increased the number of lung PD-1^{hi} ILC2s, and to a lesser extent PD-1^{hi} cNK and ILC1

cells (Fig. 4a). PD-1^{hi} ILC2s were abundant on day 3 and day 7 after infection, but much fewer at 14 or 21 days after infection when mice recovered from the infection (Fig. 4b). Therefore, PD-1 may be an indicator of ILC activation, similar to that in T and B cells²⁶. Indeed, in *Il13*^{+/-tdTomato} reporter mice²⁷, all IL-13^{hi} ILC2s were PD-1^{hi} (Fig. 4c, d).

To explore the functions of PD-1^{hi} ILCs further, we injected a PD-1 antibody (denoted as J43), which can reduce mouse CD4⁺PD-1⁺ T-cell numbers by complement-dependent cytotoxicity²⁸. Repeated administration of J43 reduced PD-1^{hi} T cells and tissue-resident ILCs in adult mice (detected by an alternative PD-1 antibody RMP1-30) (Extended Data Fig. 10c). The failure to detect PD-1 was not due to epitope competition, as J43 and RMP1-30 effectively co-stained lung ILC2s (Extended Data Fig. 10d). We next pre-conditioned *Rag1*^{-/-} mice with J43 before influenza infection. 7 days after infection, J43 injection caused reduction of total lung ILC2 (Bcl11b⁺) (Fig. 4e, f), with almost complete loss of PD-1^{hi} ILC2 cells (Fig. 4g, h) and IL-13-producing ILC2s (Fig. 4i). Moreover, type 2 cytokines were sparsely detected in bronchoalveolar lavage fluid (BALF) of J43-treated mice (Fig. 4j). Similarly, PD-1^{hi} ILC1s, cNKs and ILC3s were also depleted with reduction of type 1 and 3 cytokines (Extended Data Fig. 10e, f). Anti-PD-1 treatment may thus provide an effective approach to modulate ILC responses.

The protease allergen papain causes occupational asthma and airway inflammation by activating potent ILC2 cytokine response^{29,30}.

We found that papain challenge increased lung PD-1^{hi} ILC2 cells in the *Rag1*^{-/-} mice (Extended Data Fig. 10g). Unlike reduction by ILC2 depletion with CD25 antibody³⁰, J43 treatment specifically depleted PD-1^{hi} and IL-5-producing ILC2 cells (Extended Data Fig. 10h), and resulted in less eosinophil accumulation in BALF (Fig. 4k, l). Histological examination of lung sections revealed that J43 reduced lung eosinophil cells (Fig. 4m). Periodic-acid–Schiff staining demonstrated mucus production inhibition (Fig. 4n). Depleting PD-1^{hi} ILCs thus effectively prevented lung inflammation caused by papain.

scRNA-seq of bone marrow ILCs thus enabled identification of distinct ILC progenitor subsets, delineation of ILC development, and discovery of PD-1 as a marker of innate lymphoid progenitors and effector ILCs. These results present new therapeutic opportunities for manipulation of ILCs to achieve optimal immune responses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 22 September 2016.

Published online 29 September 2016.

1. Eberl, G., Colonna, M., Di Santo, J. P. & McKenzie, A. N. Innate lymphoid cells. Innate lymphoid cells: a new paradigm in immunology. *Science* **348**, aaa6566 (2015).
2. Artis, D. & Spits, H. The biology of innate lymphoid cells. *Nature* **517**, 293–301 (2015).
3. Chea, S. *et al.* Single-cell gene expression analyses reveal heterogeneous responsiveness of fetal innate lymphoid progenitors to Notch signaling. *Cell Reports* **14**, 1500–1516 (2016).
4. Constantinides, M. G., McDonald, B. D., Verhoef, P. A. & Bendelac, A. A committed precursor to innate lymphoid cells. *Nature* **508**, 397–401 (2014).
5. Klose, C. S. *et al.* Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. *Cell* **157**, 340–356 (2014).
6. Yang, Q. *et al.* TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. *Nat. Immunol.* **16**, 1044–1050 (2015).
7. Yu, X. *et al.* The basic leucine zipper transcription factor NFIL3 directs the development of a common innate lymphoid cell precursor. *eLife* **3**, (2014).
8. Ishizuka, I. E. *et al.* Single-cell analysis defines the divergence between the innate lymphoid cell lineage and lymphoid tissue-inducer cell lineage. *Nat. Immunol.* **17**, 269–276 (2016).
9. Spits, H. *et al.* Innate lymphoid cells—a proposal for uniform nomenclature. *Nat. Rev. Immunol.* **13**, 145–149 (2013).
10. Serafini, N., Voshchenrich, C. A. & Di Santo, J. P. Transcriptional regulation of innate lymphoid cell fate. *Nat. Rev. Immunol.* **15**, 415–428 (2015).
11. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
12. Yu, Y. *et al.* Bcl11a is essential for lymphoid development and negatively regulates p53. *J. Exp. Med.* **209**, 2467–2483 (2012).
13. Xu, W. *et al.* NFIL3 orchestrates the emergence of common helper innate lymphoid cell precursors. *Cell Reports* **10**, 2043–2054 (2015).
14. Seillet, C. *et al.* Nfil3 is required for the development of all innate lymphoid cell subsets. *J. Exp. Med.* **211**, 1733–1740 (2014).
15. Seehus, C. R. *et al.* The development of innate lymphoid cells requires TOX-dependent generation of a common innate lymphoid cell progenitor. *Nat. Immunol.* **16**, 599–608 (2015).
16. Zook, E. C. *et al.* The ETS1 transcription factor is required for the development and cytokine-induced expansion of ILC2. *J. Exp. Med.* **213**, 687–696 (2016).
17. Weber, B. N. *et al.* A critical role for TCF-1 in T-lineage specification and differentiation. *Nature* **476**, 63–68 (2011).
18. McConnell, M. J. *et al.* Growth suppression by acute promyelocytic leukemia-associated protein PLZF is mediated by repression of c-myc expression. *Mol. Cell. Biol.* **23**, 9375–9388 (2003).

19. Hoyler, T. *et al.* The transcription factor GATA-3 controls cell fate and maintenance of type 2 innate lymphoid cells. *Immunity* **37**, 634–648 (2012).
20. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12**, 252–264 (2012).
21. Yagi, R. *et al.* The transcription factor GATA3 is critical for the development of all IL-7R α -expressing innate lymphoid cells. *Immunity* **40**, 378–388 (2014).
22. Yu, Y. *et al.* The transcription factor Bcl11b is specifically expressed in group 2 innate lymphoid cells and is essential for their development. *J. Exp. Med.* **212**, 865–874 (2015).
23. Walker, J. A. *et al.* Bcl11b is essential for group 2 innate lymphoid cell development. *J. Exp. Med.* **212**, 875–882 (2015).
24. Huang, Y. *et al.* IL-25-responsive, lineage-negative KLRG1^{hi} cells are multipotential ‘inflammatory’ type 2 innate lymphoid cells. *Nat. Immunol.* **16**, 161–169 (2015).
25. Sharma, P. & Allison, J. P. Immune checkpoint targeting in cancer therapy: toward combination strategies with curative potential. *Cell* **161**, 205–214 (2015).
26. Agata, Y. *et al.* Expression of the PD-1 antigen on the surface of stimulated mouse T and B lymphocytes. *Int. Immunol.* **8**, 765–772 (1996).
27. Barlow, J. L. *et al.* Innate IL-13-producing nuocytes arise during allergic lung inflammation and contribute to airways hyperreactivity. *J. Allergy Clin. Immunol.* **129**, 191–198 (2012).
28. Kasagi, S. *et al.* Anti-programmed cell death 1 antibody reduces CD4⁺PD-1⁺ T cells and relieves the lupus-like nephritis of NZB/W F1 mice. *J. Immunol.* **184**, 2337–2347 (2010).
29. Novoy, H. S., Marchioli, L. E., Sokol, W. N. & Wells, I. D. Papain-induced asthma—physiological and immunological features. *J. Allergy Clin. Immunol.* **63**, 98–103 (1979).
30. Halim, T. Y., Krauss, R. H., Sun, A. C. & Takei, F. Lung natural helper cells are a critical source of Th2 cell-type cytokines in protease allergen-induced airway inflammation. *Immunity* **36**, 451–463 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Colucci and J. Di Santo for providing *Rag2*^{-/-}*Il2rg*^{-/-} mice. We thank the Sanger Institute RSF (J. Bussell, D. Key, A. Kirton, L. Bulman, S. Kemp, P. Green, P. Zielesinski, R. Lacey, C. Rogerson, A. Logan and G. Notley), Flow Cytometry Core Facility (B. L. Ng, J. Graham and C. Hall), Single Cell Genomic Core Facility (S. Loren and I. Bronner) and DNA sequencing pipeline (N. Smerdon) for technical assistances. We thank K. Chen, J. Pramanik and R. Miragaia for technical help. C.W. is supported by the Plan of Youth Growth from Shanghai Municipal Agricultural Committee (Hunongqingzi (2015. No. A-35)). L.L. is funded by National Natural Science Foundation of China (31370904, 81671579). G.T.B. is supported by the Australian Research Council (Future Fellowship FT110100283) and the National Health and Medical Research Council (Fellowship 10402092). A.N.J.M. is supported by the Medical Research Council (U105178805) and Wellcome Trust (100963/Z/13/Z). This work is supported by Wellcome Trust (grant number 098051) (P.L.).

Author Contributions Y.Y. designed research, performed experiments and analysed data. J.C.H.T. performed all of the bioinformatics analyses. C.W. performed experiments. S.C. and C.B. performed influenza infection. J.W. generated the *Bcl11b*^{tdTomato} conditional knockout reporter mice. X.C. performed ChIP–PCR. L.K. did the histologic section staining. L.S.C. analysed the histologic data. L.L. contributed intellectually to the PD-1 experiments. G.T.B. and A.N.J.M. provided *Id2*^{GFP} and *Il13*^{+/tdTomato} reporter mice, respectively. S.A.T. and G.D. provided intellectual input for the experiments performed in their laboratories. Y.Y., J.C.H.T. and P. L. wrote the paper. Y. Y. and P. L. conceived the PD-1 as an ILC marker concept. P. L. supervised the research.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P. L. (pl2@sanger.ac.uk).

Reviewer Information *Nature* thanks I. Amit and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Mice. The *Bcl11b*^{tdTomato} conditional knockout reporter mice were generated on C57BL/6 background (Extended Data Fig. 6b). *Vav*-Cre (018968) and *Zbtb16*^{GFP} (024529) mice were purchased from The Jackson Laboratory. *Id2*^{GFP} mice were from a C57BL/6 background³¹. *Il-13*^{+tdTomato} mice were from a Balb/c background²⁷. *Rag2*^{-/-}*Il2rg*^{-/-} mice (C57BL/6 CD45.1) were used as recipients for adoptive transfers³². Mice aged 6–12 weeks were used for all experiments. Both sexes were included without randomization. No statistical method was used to predetermine sample size. The number of mice used in each experiment to reach statistical significance was determined on basis of preliminary data. Littermate controls were used whenever possible. All mice used were maintained at RSF of the Sanger Institute. Housing and breeding of mice and experimental procedures were carried out according to the UK 1986 Animals Scientific Procedure Act and Animal Welfare and Ethical Review Body (AWERB) of the Wellcome Trust Sanger Institute. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Reagents. Fluorochrome- or biotin-labelled monoclonal antibodies (clones denoted in parenthesis) against B220 (RA3-6B2), CD19 (6D5), CD3ε (145-2C11), CD5 (53-7.3), CD8α (53-6.7), TCRβ (B20.6), TCRγδ (GL3), NK1.1 (PK136), Nkp46 (29A1.4), CD49a (HMα1), CD49b (DX5), CD11b (M1/70), CD11c (N418), Gr1 (RB6-8C5), Ter119 (TER-119), c-kit (2B8), Flt3 (A2F10), Sca1 (D7), IL-7Rα (SB/199), CD25 (PC61), CD45.1 (A20), CD45.2 (104), CD45 (30-F11), CD244 (2B4), α₄β₇ (DATK32), IL-33R (RWST2-2), CD122 (TM-b1), CD94 (18D3), CD226 (10E5), CXCR3 (CXCR3-173), CXCR5 (2G8), CCR9 (CW-1.2), CCR6 (29-2L17), PD-1 (29F.1A12, J43 or RMP1-30), Gata3 (L50-823), RORγt (Q31-378), PLZF (R17-809) and IL-5 (TRFK5) were purchased from BD Biosciences, Biolegend or eBioscience. TCF-1 (C63D9) and TOX (REA473) were purchased from Cell Signaling Technology and Miltenyi Biotec, respectively. LEGENDplex Mouse Th Panel (Biolegend, 740005) was used for BALF cytokines measurement. The PD-1 antibody (J43, BE0033-2) and isotype control antibody (Armenian Hamster IgG, BE0091) were purchased from Bio X cell for *in vivo* administration.

Flow cytometry and cell sorting. Red blood cells were removed using ACK lysis buffer (Lonza, 10-548E). Cells were suspended in a solution of 2% (vol/vol) FBS in PBS. Fc receptors were blocked with anti-CD16 (2.4G2) before antibody labelling. Cells were stained with antibodies on ice for 20 min before washing. Intracellular staining was performed according to the instruction of the FOXP3 Fix/Perm Buffer Set (Biolegend, 421403). Cells were analysed on the LSRFortessa cell analyser (BD) or sorted on the MoFlo XDP cell sorter (Beckman Coulter) according to the manufacturers' standard operating procedures, respectively. Data were analysed with FlowJo version 10.0.7 software (Tree Star).

Preparation of cell suspensions. Bone marrow cells were isolated by gently crushing femurs and/or tibias before filtration (70-μm filter). Cells from lung and small intestine lamina propria were prepared according to the instructions of Lung Dissociation Kit (Miltenyi Biotec, 130-095-927) and Lamina Propria Dissociation Kit (Miltenyi Biotec, 130-097-410), respectively. The tissues were digested in a shaking water bath at 37°C for 15–20 min.

Adoptive transfers *in vivo*. Highly purified cell populations by flow cytometry were injected intravenously into sublethally irradiated (1 × 450 rads) *Rag2*^{-/-}*Il2rg*^{-/-} recipient mice (CD45.1⁺) via the tail vein. The drinking water was supplemented with antibiotics for 2 weeks after irradiation.

***In vitro* culture assay.** PD-1^{hi}, PD-1^{hi}Bcl11b⁺, PD-1^{hi}Bcl11b⁻, PD-1^{hi}IL-25R⁺ and PD-1^{hi}IL-25R⁻ cells were individually sorted into 96-well plates containing Mitomycin C (10 μg ml⁻¹)-treated OP9 stromal cells (70% confluent). Cells were cultured in RPMI160 medium (10% FBS and 80 μM 2-Mercaptoethanol) supplemented with 20 ng ml⁻¹ IL-7 (PeproTech) and 50 ng ml⁻¹ mSCF (PeproTech). All of the cells were analysed on days 10–12 of culture. For *Il17rb* overexpression rescue experiment, transduced *Bcl11b*-deficient PD-1^{hi} cells were co-cultured with ILC2P (CD45.1) in the presence of 20 ng ml⁻¹ IL-7, 20 ng ml⁻¹ SCF and 20 ng ml⁻¹ IL-33 for 14 days.

Retroviral transduction of ILC precursors. The Phoenix retroviral packaging system was used with the transfection reagent FuGENE 6 (Promega). Retroviral supernatants were centrifuged (2,000g) at 32°C for 90 min twice on 50 μg ml⁻¹ RetroNectin (TaKaRa) precoated 24-well plates according to the manufacturer's instruction. PD-1^{hi}ILC progenitors were transduced by centrifuge at 500g for 30 min. Mouse *Il17rb* cDNA was bought from Origene (MC205843) and was cloned into pMXs retroviral vector. The empty retroviral vector carrying GFP only was used as the control.

Cell lines. OP9 and OP9-DL1 cells were kindly provided by J. C. Zúñiga-Pflücker (University of Toronto). Cells were maintained at 37°C in a humidified 5% CO₂ atmosphere.

Influenza infection. The mice were anaesthetized with 3% isoflurane and were inoculated intranasally with influenza A virus X31 (H3N2) in 50 μl PBS of 10⁴

plaque-forming units. The virus was grown and collected from embryonated chicken eggs (48–72 h). At day 5 after infection, lung and BALF (lavage with 0.4 ml PBS) were collected.

Papain administration. The mice were anaesthetized with 3% isoflurane and then were intranasally administered with 30 μg of papain (Acros Organics) in 30 μl PBS every day for 5 days. Lungs and BALF were collected and analysed at day 6.

Lung histologic sections. The left lobes of the lungs were fixed with 4% paraformaldehyde and embedded in paraffin. 5 μm sections were used for staining with haematoxylin and eosin or periodic-acid-Schiff stain. Image acquired with a Leica DM4000B LED microscope, an Olympus DP72 camera, and cellSens version 1.4 imaging software. Only the brightness and contrast of the whole image were adjusted by the Adobe photoshop (CS5, Adobe systems Inc.).

Generation of single-cell RNA-seq library. The single-cell mRNA-seq library was generated following the SMART-seq2 protocol described in ref. 33. In short, single bone marrow progenitors were sorted into 96-well plates pre-filled with lysis buffer and external RNA spike-ins (Ambion) (1:500,000). First-strand synthesis and template-switching were then performed, followed by 25-cycle of pre-amplification. Complementary DNAs were purified by AMPure XP magnetic beads (Agencourt) using an automated robotic workstation (Zephyr). Quality of cDNAs was checked with the Bioanalyzer (Agilent) using high sensitivity DNA chip. Multiplex (96-plex) libraries were constructed and amplified using Nextera XT library preparation kit (Illumina). The libraries were then pooled and purified with AMPure XP magnetic beads. The quality of the library is then assessed by the Bioanalyzer (Agilent) before submission to the DNA sequencing pipeline at the Wellcome Trust Sanger Institute. Pair-ended 75-bp reads were generated by HiSeq2000 sequencers.

Read alignment and gene quantification. The pre-processed BAM files with the same indexes were first merged and converted to raw FASTQ files. The FASTQ sequences were then realigned to a modified GRCm38 mouse genome with the sequences of the 92 ERCC spike-ins added. The alignment was performed using STAR³⁴ and sorted for expression quantification by HTSeq-count³⁵. The counting was performed with a modified gene annotation file (Ensembl GRCm38.75 + 92 ERCC spike-ins) with parameter '-s no' in default union mode. Cells with unique counts less than 500,000 or gene detected less than 2500 or count mapping to mitochondrial more than 10% were removed. The count matrix was then normalized by the size factors calculated using the external ERCC spike-ins by DESeq2 (ref. 36). For the Lin⁻PD-1^{hi} cell single-cell mRNA-seq dataset, cells with unique counts less than 500,000 or count mapping to mitochondrial more than 10% were removed.

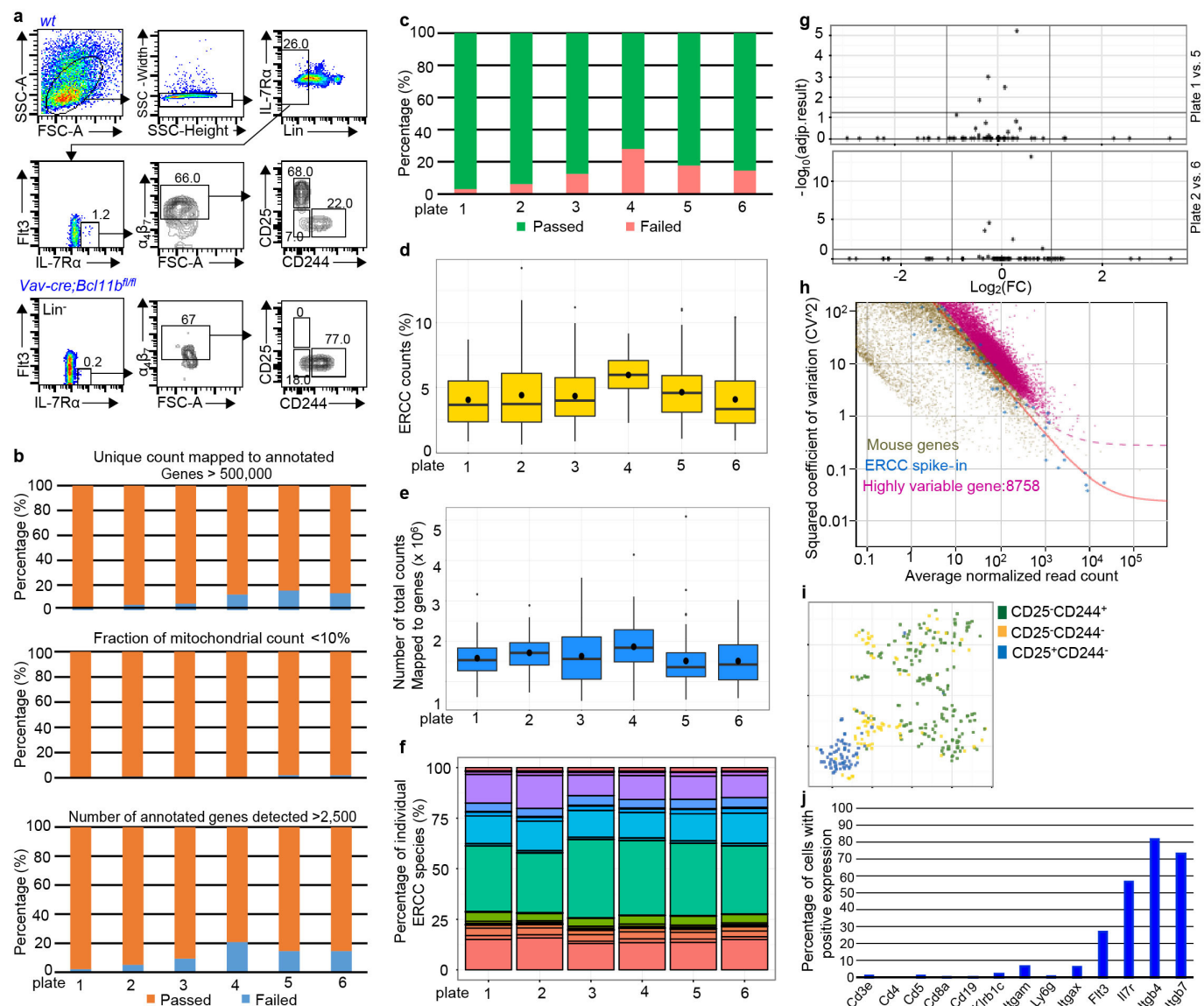
Identification of highly variable genes and cell clustering. The details of the statistical model for testing highly variable genes have been described in ref. 11. The minimal biological dispersion parameter was set at 0.5 and genes with *P* value less than 0.1 were classified as highly variable. Cell clustering was performed using the 'Rtsne' package. The count matrix containing both wild type and mutant ILCs were first normalized by the size factors calculated using their corresponding ERCC count matrix. The matrix was then log₂-transformed and only genes that are highly variable were included as input for clustering. The Rtsne package is an R wrapper of the Barnes-Hut t-SNE C++ implementation of van der Maaten and Hinton³⁷. The seed was set at 10 for reproducibility and perplexity set at 10 and 2 for the initial and Lin⁻PD-1^{hi} dataset, respectively. Cell clusters were then visualized in the biaxial scattered plot and grouped into distinct clusters on the basis of their spatial proximity and specific expression pattern of ILC lineage regulators. Specific genes of a particular cluster were defined as genes showing normalized count expression higher than that of the other clusters by one fold and expression greater than 100 normalized counts.

Gene set enrichment analysis. To identify gene signature enrichment between different cell clusters, the microarray gene expression data of different types of innate lymphoid cells were downloaded from Robinette *et al.*³⁸. The data were normalized using the *Affy* package in R. The normalized probe matrix was then converted into gene names using Biomart. The expression of specific genes was calculated as the average of the normalized signal of the different probes of the same genes. Specific gene sets of different type of innate lymphoid cells (ILC1/NK, ILC2 and ILC3/LTi) were defined as genes that showed higher average expression (>1 normalized expression unit) than the other two types of ILCs. The enrichment scores were then calculated using the javaGSEA application (version 2.2.1) available online (<http://www.broadinstitute.org/gsea/downloads.jsp>) with default settings³⁹.

Statistical analysis. All the experiments were not blinded to allocation during experiments. The outcome assessment of the experiments corresponding to Fig. 4m, n was analysed in a single-blinded manner. All other outcome assessments were not blinded. The statistical analysis was conducted with Microsoft Excel or Prism 6 (GraphPad). *P* values were calculated using a two-tailed Student's *t*-test.

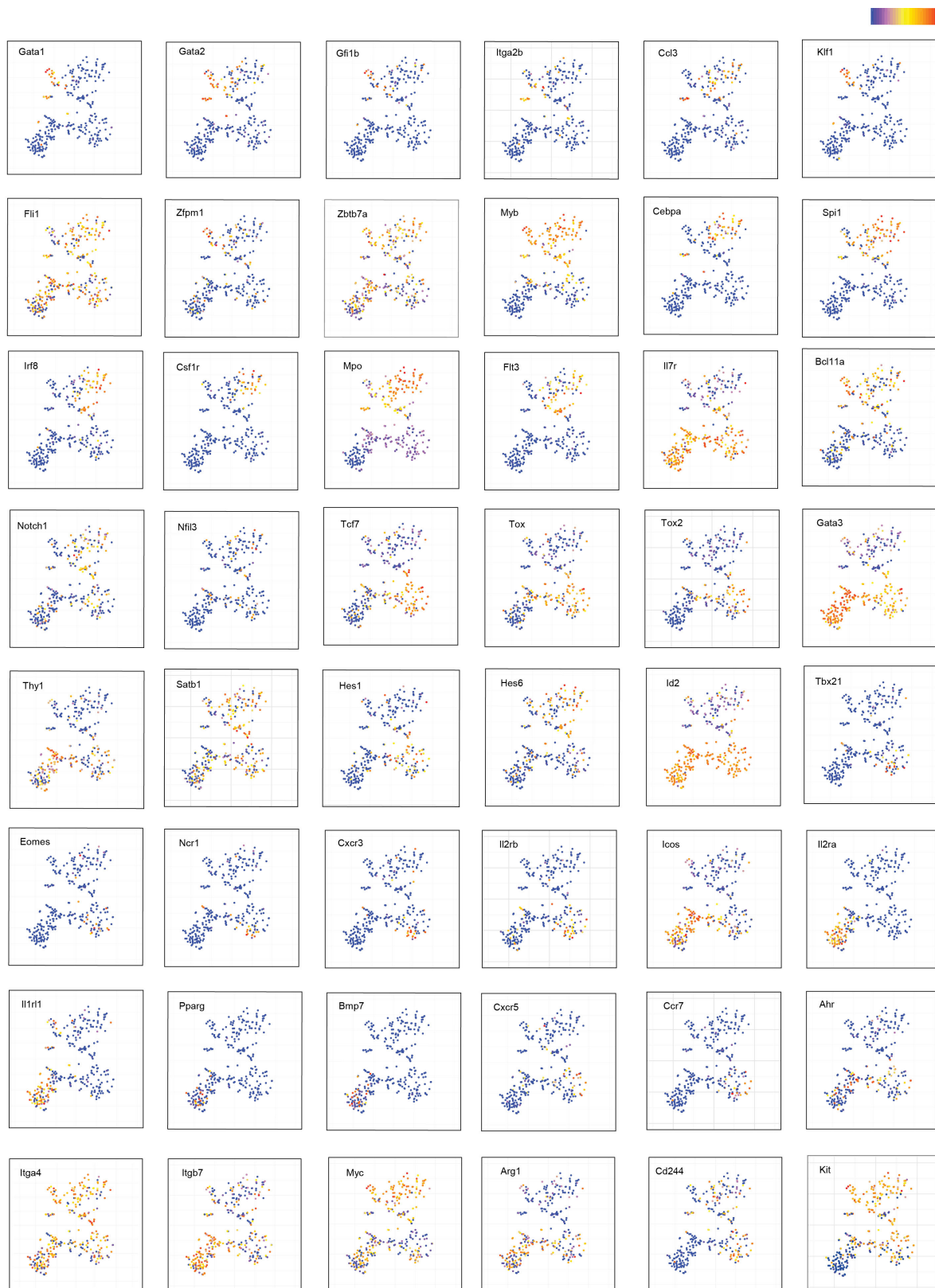
Data availability. RNA-seq data that support the findings of this study have been deposited in the European Nucleotide Archive database and are accessible through the accession number ERP011804. All other relevant data are available from the corresponding author on request.

31. Jackson, J. T. *et al.* Id2 expression delineates differential checkpoints in the genetic program of CD8 α ⁺ and CD103⁺ dendritic cell lineages. *EMBO J.* **30**, 2690–2704 (2011).
32. Serafini, N. *et al.* *Gata3* drives development of ROR γ t⁺ group 3 innate lymphoid cells. *J. Exp. Med.* **211**, 199–208 (2014).
33. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protocols* **9**, 171–181 (2014).
34. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
35. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
37. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **4**, 1–48 (2008).
38. Robinette, M. L. *et al.* Transcriptional programs define molecular characteristics of innate lymphoid cell classes and subsets. *Nat. Immunol.* **16**, 306–317 (2015).
39. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

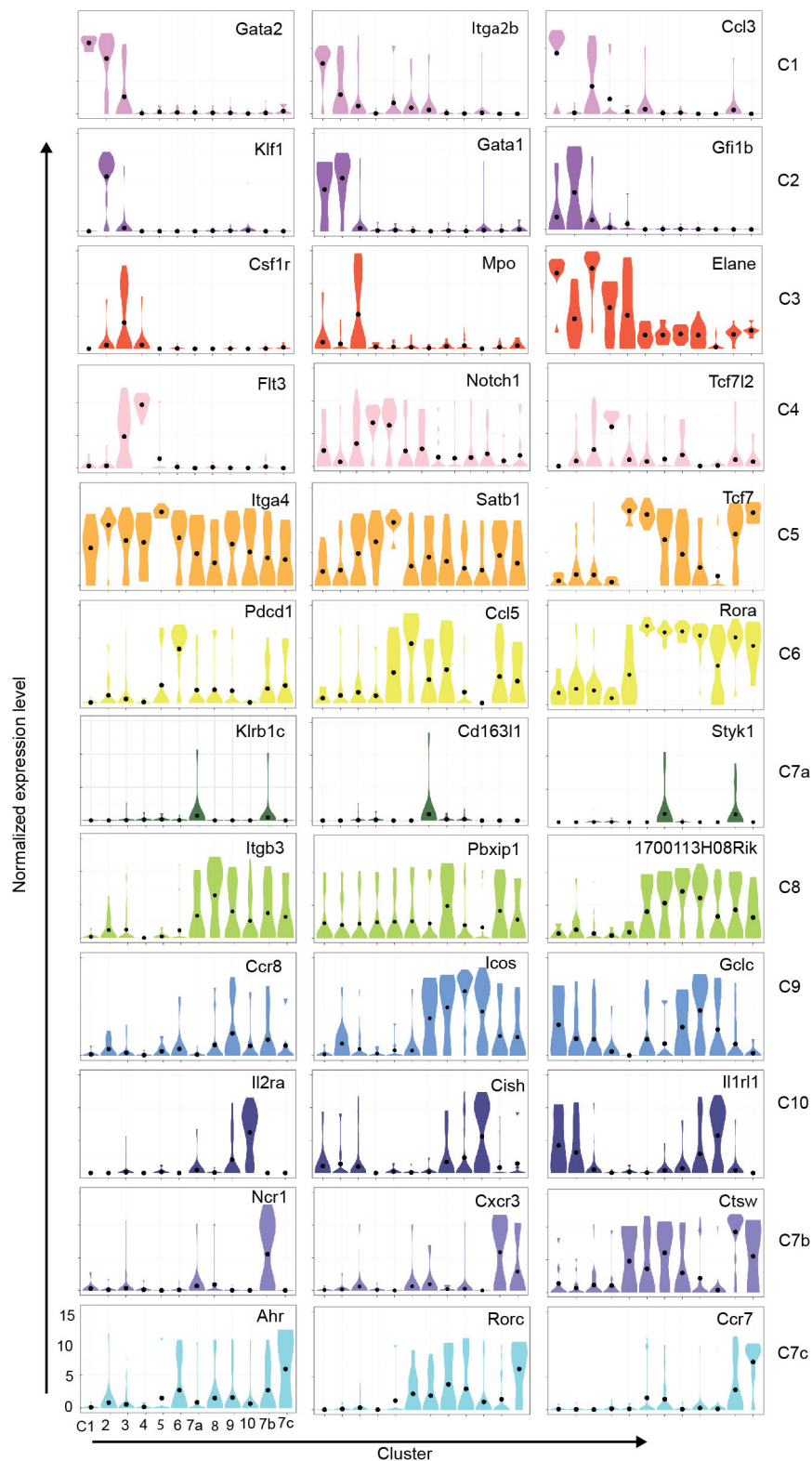


Extended Data Figure 1 | Quality control of scRNA-seq data of ILC-progenitor-enriched bone marrow cells. **a**, FACS sorting strategies of the adult bone marrow cells from wild-type or *Vav-Cre-Bcl11b^{fl/fl}* mice. Flt3^{lo} or IL-7Rα^{lo} cells were included to detect more ILC progenitors. Lin⁻Flt3^{lo}-IL-7Rα^{lo/+}α₄β₇⁺ cells were further divided into three populations (CD244⁺CD25⁻, CD244⁻CD25⁻ and CD244⁻CD25⁺). We sorted two 96-well plates of CD244⁺CD25⁻, one plate of CD244⁻CD25⁻ and one plate of ILC2 progenitors CD244⁻CD25⁺ to include most ILC progenitors for scRNA-seq. Two 96-well plates of Lin⁻Flt3⁻IL-7Rα⁺α₄β₇⁺ bone marrow cells from *Vav-Cre-Bcl11b^{fl/fl}* mice were purified to investigate early ILC2 development defects. Lin: CD19, CD3, CD4, CD5, CD8, TCRβ, TCRγδ, NK1.1, CD11b, Gr-1, CD11c and Ter119. **b**, Column charts show the fraction of cells passing specific quality control criteria in each plate: unique count mapped to annotated genes >500,000 (top panel); count mapped to mitochondrial-encoded genes <10% (middle panel); and number of annotated gene detected >2,500 (bottom panel). **c**, Percentage of cells passing all criteria. **d**, The percentages of ERCC RNA spike-ins in

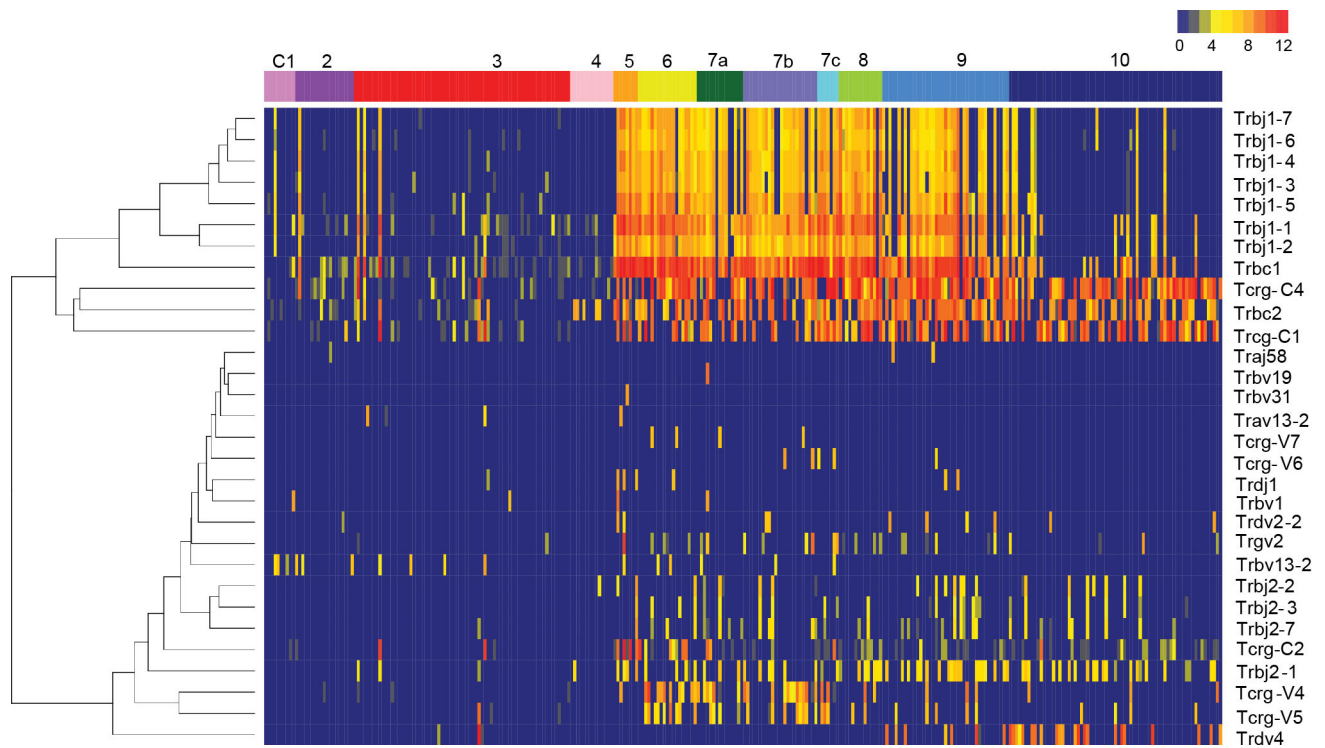
each plate. The black dots represent the mean of the dataset. **e**, The total number of unique counts mapped to annotated genes in different plates. The black dots represent the mean of the dataset. **f**, The fractions of the 92 external ERCC RNA spike-ins in different plates. **g**, Kolmogorov–Smirnov test of individual ERCC spike-ins between the two plates did not detect any ERCC spike-ins showing significantly different (\log_2 fold change > 1 , and adjusted P value < 0.05) levels. The two vertical lines mark the \log_2 fold change levels of -1 and 1 , the horizontal line marks the adjusted P value threshold of 0.05 . **h**, Identification of highly variable genes. Brown points represent annotated mouse genes. Blue points represent external ERCC RNA spike-ins. The magenta points represent the mouse genes that show significantly higher variability (false discovery rate < 0.1). The solid line represents the fit of the technical noise, the dashed line represents the 50% biological CV (coefficient of variation). **i**, Biaxial t-SNE clustering of the sequenced wild-type cells. **j**, Column chart comparing the percentage of cells which show detectable mRNA expression of lineage markers in the wild-type bone marrow cells.



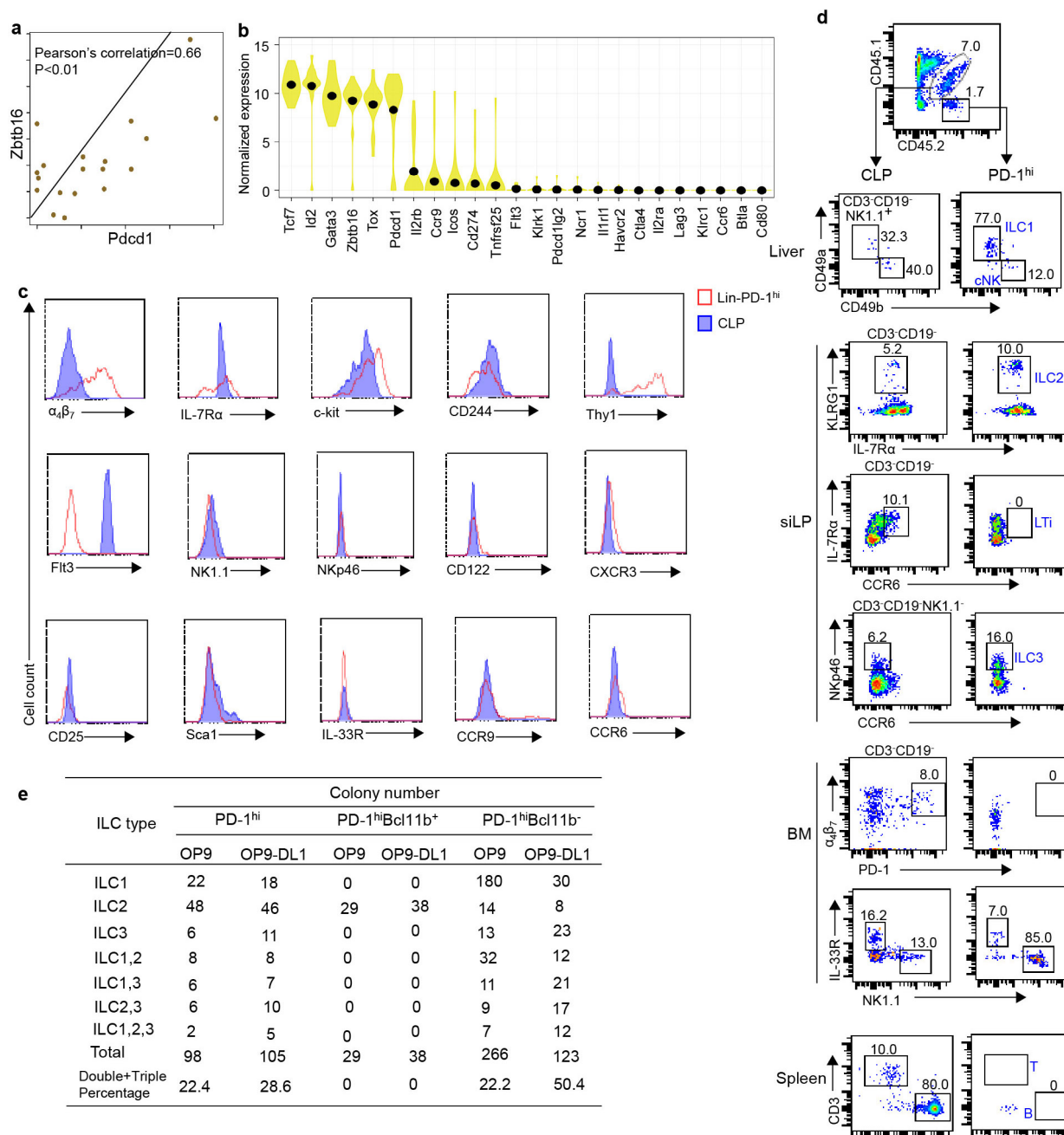
Extended Data Figure 2 | t-SNE plots showing expression and distribution of genes described in the manuscript. The colour key shows the expression level.



Extended Data Figure 3 | Violin plots of genes highly enriched in specific clusters. The y axis indicates the \log_2 (normalized count + 1) expression levels. The black point indicates the mean of expression level. The x axis indicates different clusters.



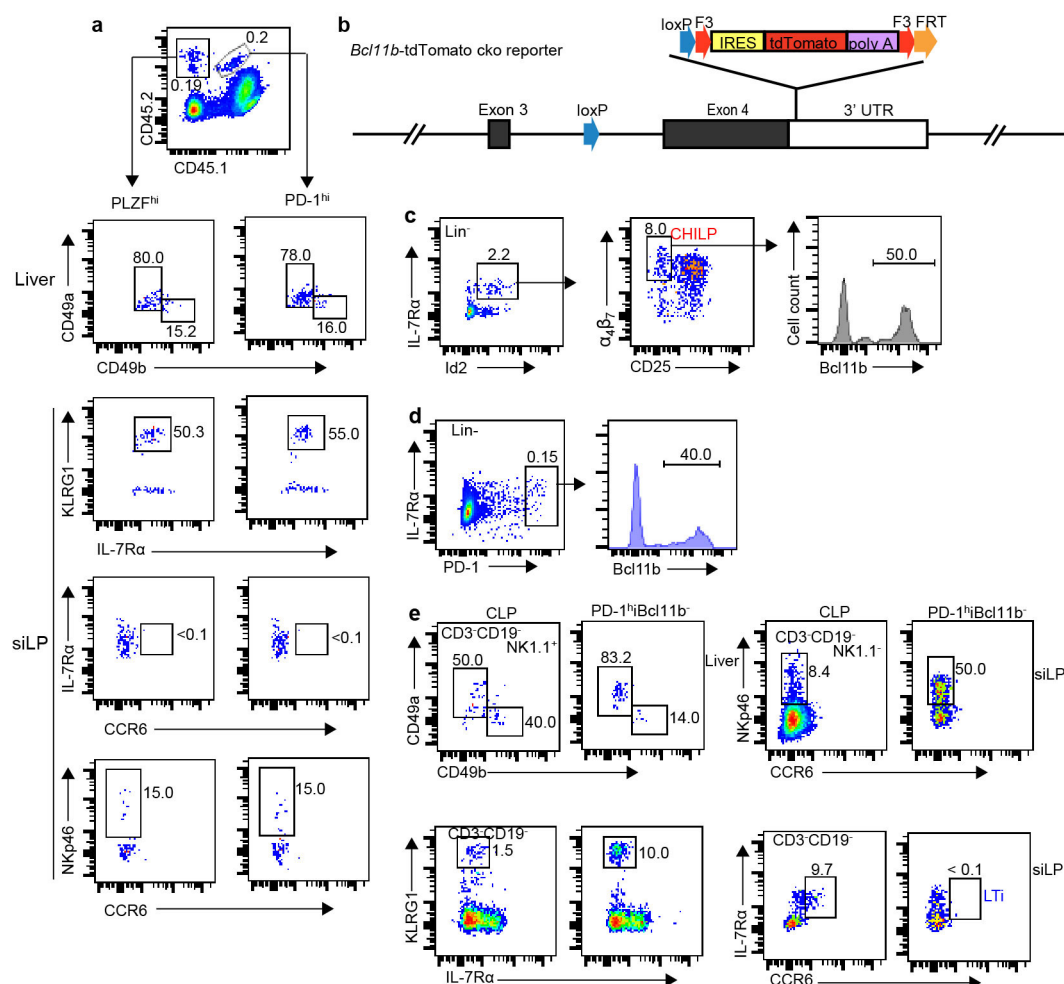
Extended Data Figure 4 | Heat map of TCR transcripts in different cell clusters.



Extended Data Figure 5 | PD-1^{hi} expression marks ILC progenitors.

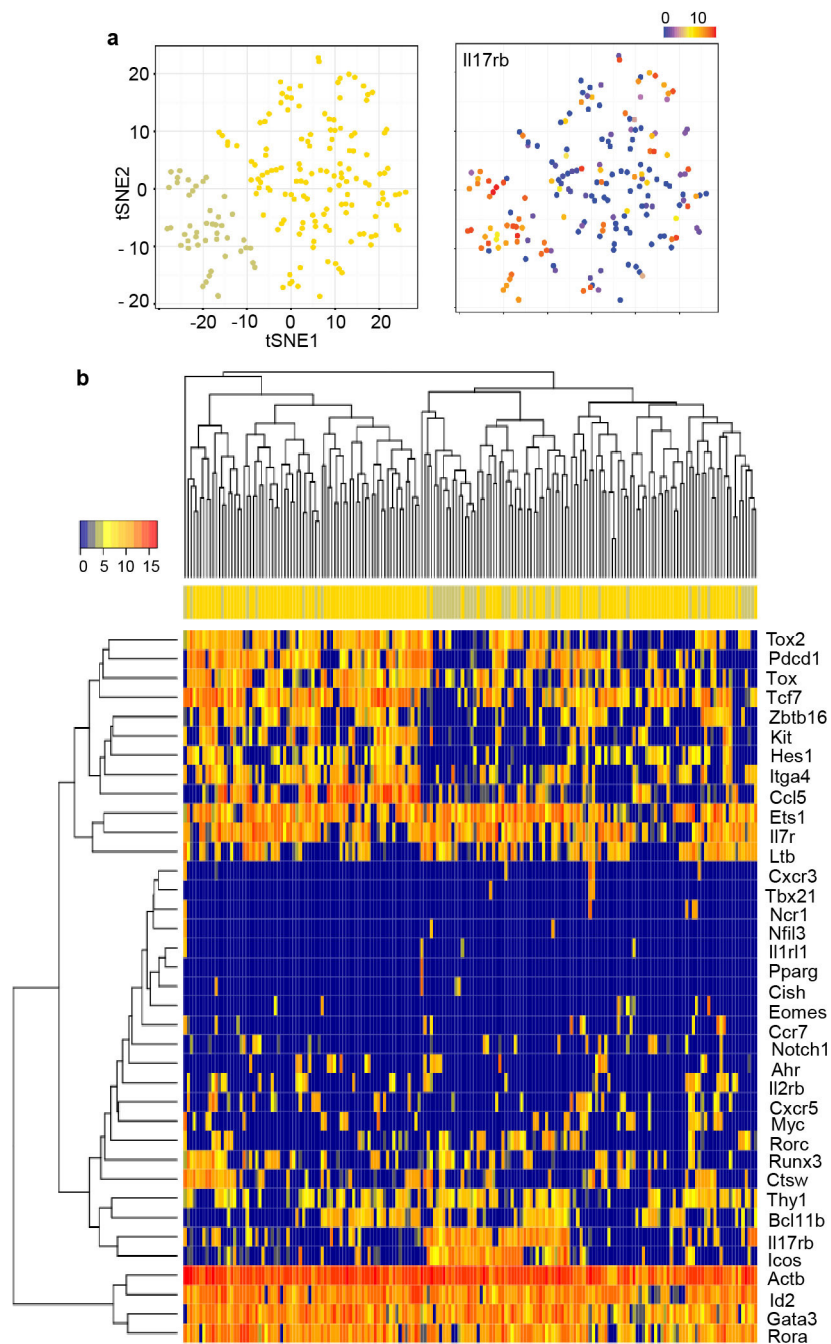
a, Correlation of expression levels of *Pdcd1* and *Zbtb16* in C6 cells. Correlation was calculated by Pearson's method. The fit represents the line of linear regression. **b**, Violin plot showing the selected gene expression in PD-1^{hi} cells of C6. The y axis indicates the log₂ (normalized count + 1) expression levels. The black point indicates the mean of expression level. **c**, Expression of ILC markers in PD-1^{hi} cells was analysed by FACS. **d**, The *in vivo* developmental potential of PD-1^{hi} cells. CD45.1 *Rag2*^{-/-} *Il2rg*^{-/-} recipients were injected with the equal numbers of CD45.1⁻CD45.2⁺

PD-1^{hi} cells and CD45.1⁺CD45.2⁺ (F₁ of CD45.1 and CD45.2 parents) common lymphoid progenitors (200–800 cells). The progenies of these donor cells were analysed by FACS after 5–7 weeks (*n* = 3 per donor cell type). **e**, Clonal analysis of PD-1^{hi} cells *in vitro*. The PD-1^{hi}, PD-1^{hi}Bcl11b⁺ and PD-1^{hi}Bcl11b⁻ bone marrow cells were FACS-purified and cultured on stromal cells and analysed by FACS. ILC1 was defined as CD45⁺NK1.1⁺Bcl11b⁻, ILC2 as CD45⁺NK1.1⁻Bcl11b⁺ and ILC3 as CD45⁺NK1.1⁻Bcl11b⁻RORγt⁺. Data are representatives of two (c) or three (d) independent experiments.

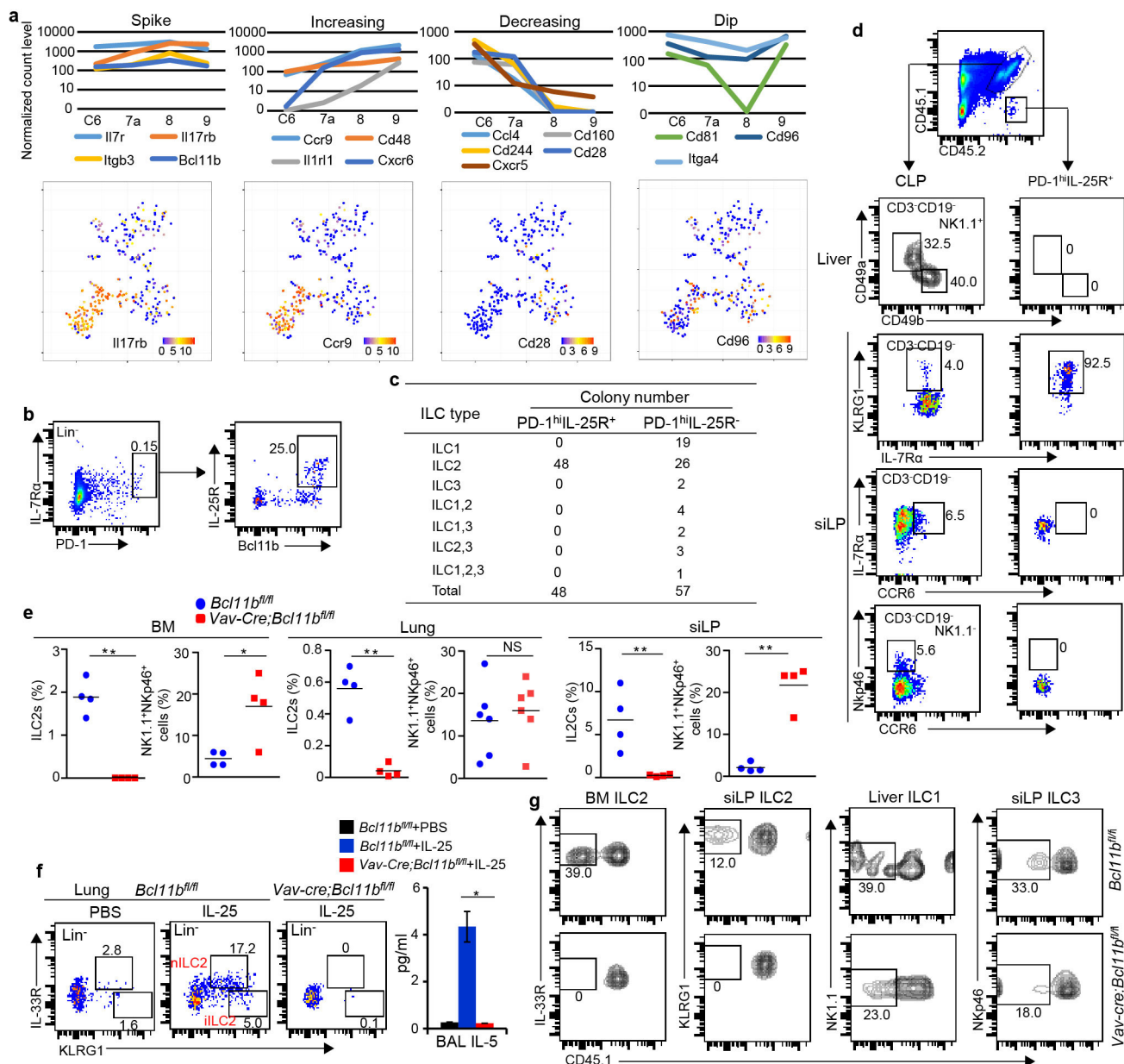


Extended Data Figure 6 | Direct comparison of PD-1^{hi} and PLZF^{hi} ILC progenitors and dissection of the heterogeneity in the ILC progenitor compartments. **a**, FACS plots show PLZF^{hi} and PD-1^{hi} cells had the same development potential *in vivo*. The equal numbers of PD-1^{hi} cells and PLZF^{hi} cells were adoptively transferred into the same recipient and analysed 3–4 weeks later ($n = 3$ per donor cell type). **b**, Schematic diagram of the *Bcl11b*^{tdTomato} conditional knockout reporter allele, where the loxP-IRES-tdTomato cassette was inserted to the 3'UTR of the *Bcl11b* locus. The other loxP site was in intron 3. Cre-loxP recombination would

delete the exon 4. The selection cassette for initial gene targeting was excised by Flpase-FRT recombination. **c**, Expression of *Bcl11b* in CHILPs were analysed in the *Id2*^{GFP};*Bcl11b*^{tdTomato} dual reporter mice ($n = 6$). **d**, Expression of *Bcl11b* in PD-1^{hi} bone marrow cells was analysed by FACS ($n = 6$). **e**, FACS analysis of the *in vivo* developmental potential of PD-1^{hi}Bcl11b⁻ cells ($n = 3$ per donor cell type). Common lymphoid progenitors were used as the donor cell control. PD-1^{hi}Bcl11b⁻ cells predominantly generated ILC1, ILC2 and ILC3. Data are representatives of two (**a**, **e**) or three (**c**, **d**) independent experiments.



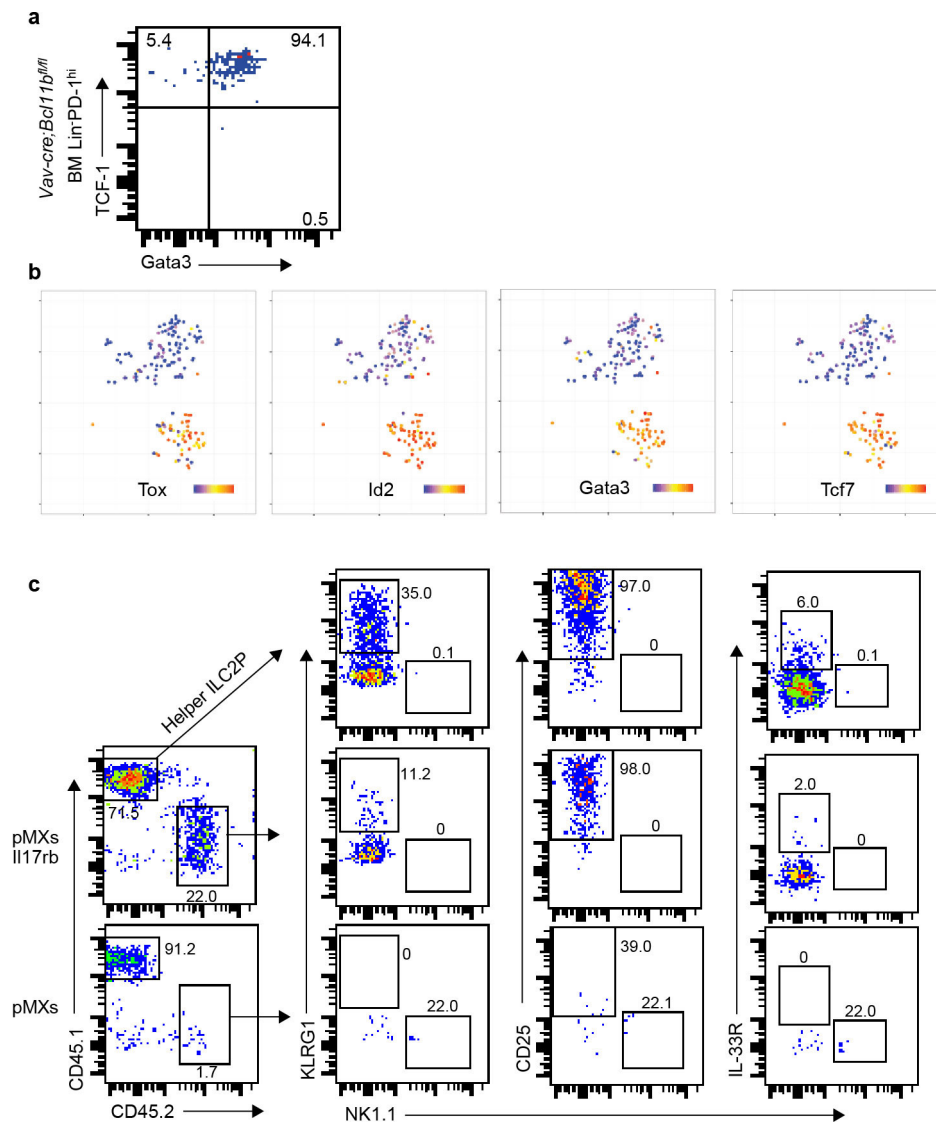
Extended Data Figure 7 | scRNA-seq analysis of PD-1^{hi} bone marrow cells. **a**, t-SNE clustering analysis of sequenced PD-1^{hi} cells detected two subpopulations. **b**, Heat map showing the hierarchical clustering result of PD-1^{hi} cells based on selected ILC regulators. The expression levels are log₂ transformed and ERCC-size factor normalized.



Extended Data Figure 8 | scRNA-seq dissection of early ILC2 development.

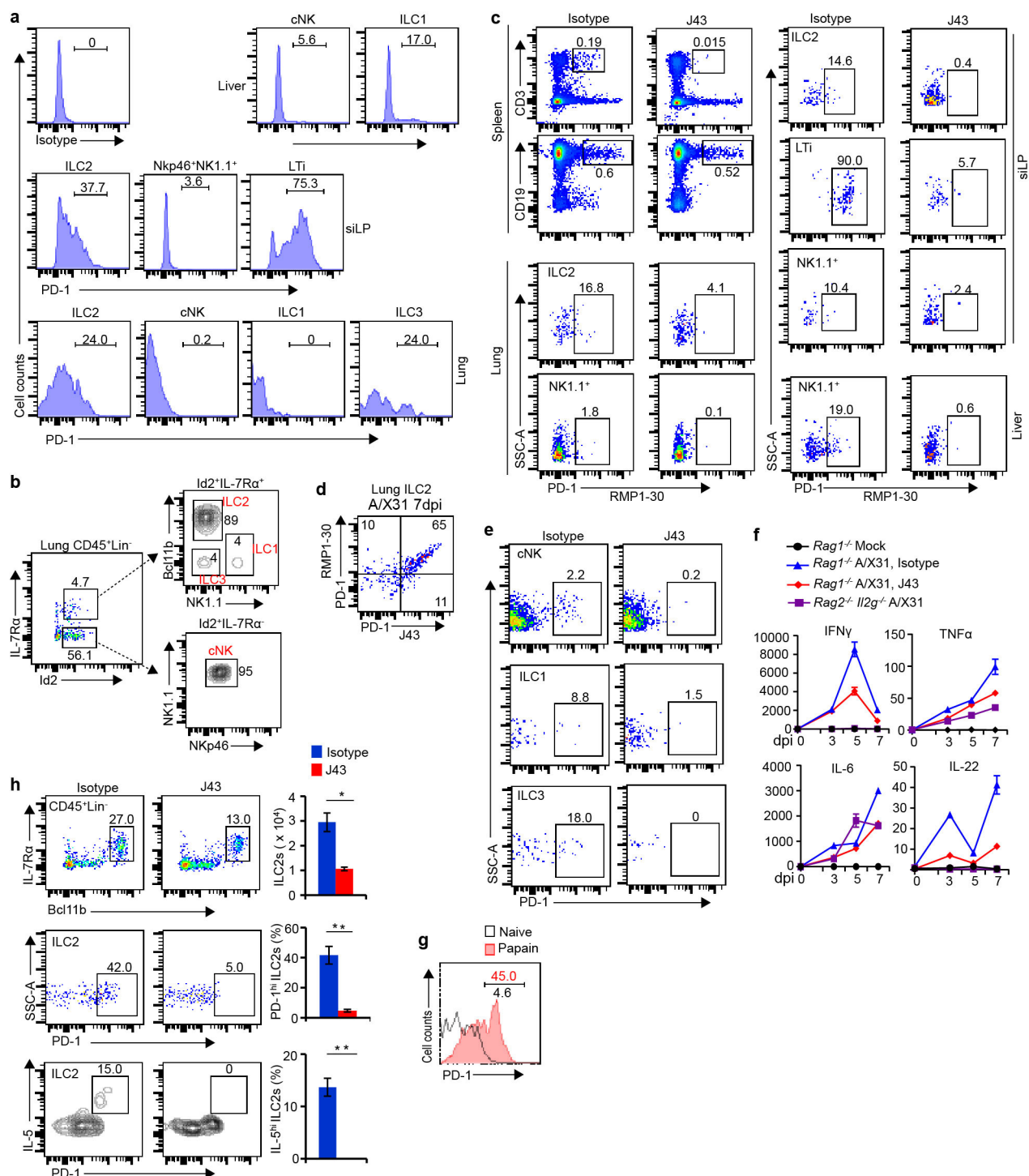
a, Analysis of scRNA-seq data identified genes showing expression changes in C6, C7a, C8 and C9 cells. Change and distribution of expression of selected genes are shown. *Il17rb* and *Bcl11b* are among the genes showing spike expression from C6 to C9, whereas *Il1r1* (IL-33R) showed steadily increased expression. The bottom t-SNE plots showing expression of representative genes. **b**, The expression of IL-25R in PD-1^{hi} bone marrow cells in the *Bcl11b^{fl/fl}* mice ($n = 3$). **c**, Clonal differentiation assay of PD-1^{hi}IL-25R⁺ and PD-1^{hi}IL-25R⁻ cells. Cells were cultured on OP9-DL1 stromal cells in the presence of IL-7 (20.0 ng ml⁻¹) and SCF (50.0 ng ml⁻¹) and were analysed 10 days later. ILC1 was defined as CD45⁺NK1.1⁺Bcl11b⁻, ILC2 as CD45⁺NK1.1⁺Bcl11b⁺ and ILC3 as CD45⁺NK1.1⁺Bcl11b⁻RORγt⁺. **d**, FACS analysis of the *in vivo* developmental potential of PD-1^{hi}IL-25R⁺ cells. CD45.1⁺*Rag2*^{-/-}*Il2rg*^{-/-} recipients were injected with equal numbers of CD45.1⁺CD45.2⁺ PD-1^{hi}IL-25R⁺ cells and CD45.1⁺CD45.2⁺ common lymphoid

progenitors (100–500 of each) and the progenies of these populations were analysed 5–7 weeks later ($n = 5$ per group). **e**, Analysis of ILCs in *Vav-Cre-Bcl11b^{fl/fl}* mice. Lin⁻IL-33R⁺IL-7Rα⁺ ILC2s, Lin⁻KLRG1⁺IL-7Rα⁺ ILC2 or Lin⁻NK1.1⁺NKp46⁺ ILCs from the bone marrow, lung or siLP were analysed ($n = 4$ per genotype), respectively. **f**, ‘Natural’ ILC2 (nILC2), ‘inflammatory’ ILC2 (iILC2) and BALF IL-5 were analysed in *Vav-Cre-Bcl11b^{fl/fl}* and the control mice after administration of IL-25 (200 ng per mouse per day) for 3 consecutive days ($n = 5$ per treated group). Error bars denote s.e.m. **g**, FACS analysis of *in vivo* developmental potential of PD-1^{hi} cells from *Vav-Cre-Bcl11b^{fl/fl}* mice. CD45.1⁺*Rag2*^{-/-}*Il2rg*^{-/-} mice were injected with CD45.2 PD-1^{hi} cells sorted from *Bcl11b^{fl/fl}* or *Vav-Cre-Bcl11b^{fl/fl}* mice. The progenies of these donor cells were analysed 4–7 weeks later by FACS ($n = 3$ per genotype). Data are representatives of three (**b**) or two (**d–g**) independent experiments. * $P < 0.05$, ** $P < 0.01$ (two-tailed *t*-test).



Extended Data Figure 9 | Restoration of development of *Bcl11b*-deficient PD-1^{hi} ILC progenitors to ILC2 by overexpressing IL-25R.
a, FACS plots showing the expression of TCF-1 and Gata3 in mutant PD-1^{hi} bone marrow cells. Protein expression was measured by intracellular antibody staining. **b**, Expression patterns of *Tox*, *Id2*, *Tcf7* and *Gata3* in the sequenced *Bcl11b*-deficient bone marrow cells.
c, Overexpressing *Il17rb* in *Bcl11b*-deficient PD-1^{hi} bone marrow cells.

The rescued cells were analysed by FACS for ILC2 surface markers. PD-1^{hi} cells sorted from *Vav-Cre-Bcl11b^{fl/fl}* mice were transduced with the *Il17rb* or control retrovirus. The infected cells were cultured on OP9-DL1 stromal cells with the helper CD45.1 ILC2 progenitors in the presence of IL-25 (20.0 ng ml⁻¹), IL-7 (20.0 ng ml⁻¹) and SCF (50.0 ng ml⁻¹). The cells were collected and analysed after two weeks of culture. Data are representatives of two (**a**, **c**) independent experiments.



Extended Data Figure 10 | PD-1^{hi} marks effector ILCs. **a**, FACS analysis of PD-1 expression on peripheral ILCs in steady-state mice ($n = 3$). **b**, Gating strategies of lung ILCs. Lung cNK cells were gated as Lin⁻Id2⁺IL-7R α ⁺NK1.1⁺NKp46⁺; lung ILC1s as Lin⁻Id2⁺IL-7R α ⁺NK1.1⁺Bcl11b⁻; lung ILC2s as Lin⁻Id2⁺IL-7R α ⁺NK1.1⁺Bcl11b⁺; and lung ILC3s as Lin⁻Id2⁺IL-7R α ⁺NK1.1⁺Bcl11b⁻. The data were from influenza-infected mice at 5 days after infection. cNKs count for at least half of the Lin⁻ leukocytes in these mice ($n = 3$). **c**, FACS analysis of PD-1 expression on CD3⁺ T cells, CD19⁺ B cells and peripheral ILCs after J43 treatment ($n = 3$). The tissues were collected at day 14 after J43 treatment. **d**, FACS plot shows the recognition of different epitopes of PD-1 by PD-1 antibody clones RMP1-30 and J43. The majority of lung PD-1^{hi} ILC2s were stained with both RMP1-30 and J43. **e**, FACS analysis of lung PD-1^{hi} cNK, ILC1 and ILC3 at 7 days after infection ($n = 3$). **f**, BALF

cytokines were quantitated as shown ($n = 3$ per group per time point). The four experimental groups were: Rag1^{-/-} mice with mock infection; Rag1^{-/-} mice infected with A/X-31 and treated with either an antibody isotype control or J43; Rag2^{-/-}Il2rg^{-/-} mice infected with A/X-31 as the ILC-deficient control. **g**, More PD-1^{hi} cells were found after papain challenge (day 6) in Rag1^{-/-} mice ($n = 3$). **h**, Rag1^{-/-} mice were pretreated with PD-1 antibody J43 or the isotype control antibody for 3 days and then administrated with papain (intranasally) for 5 consecutive days. The lung tissue was collected at day 6 for analysis ($n = 5$ per treatment). Lung ILC2s were reduced in J43 treated mice. PD-1^{hi} and IL-5-producing ILC2 were undetectable after J43 administration. Data are representatives of two (a–h) independent experiments. Error bars (f, h) denote s.e.m. * $P < 0.05$, ** $P < 0.01$ (two-tailed t -test).

On-target efficacy of a HIF-2 α antagonist in preclinical kidney cancer models

Hyejin Cho¹, Xinlin Du², James P. Rizzi², Ella Liberzon¹, Abhishek A. Chakraborty¹, Wenhua Gao¹, Ingrid Carvo^{1,3}, Sabina Signoretti^{1,3}, Richard K. Bruick⁴, John A. Josey², Eli M. Wallace² & William G. Kaelin Jr^{1,5}

Clear cell renal cell carcinoma, the most common form of kidney cancer, is usually linked to inactivation of the pVHL tumour suppressor protein and consequent accumulation of the HIF-2 α transcription factor (also known as EPAS1)¹. Here we show that a small molecule (PT2399) that directly inhibits HIF-2 α causes tumour regression in preclinical mouse models of primary and metastatic pVHL-defective clear cell renal cell carcinoma in an on-target fashion. pVHL-defective clear cell renal cell carcinoma cell

lines display unexpectedly variable sensitivity to PT2399, however, suggesting the need for predictive biomarkers to be developed to use this approach optimally in the clinic.

HIF-2 α , as a basic helix-loop-helix (bHLH)-Per/Arnt/Sim (PAS) domain protein, would usually be deemed undruggable. However, medicinal chemistry efforts have led to the development of drug-like chemicals such as PT2399 (Fig. 1a) that can bind directly to the HIF-2 α PAS B domain (Fig. 1a and Extended Data Fig. 1a, b) and prevent

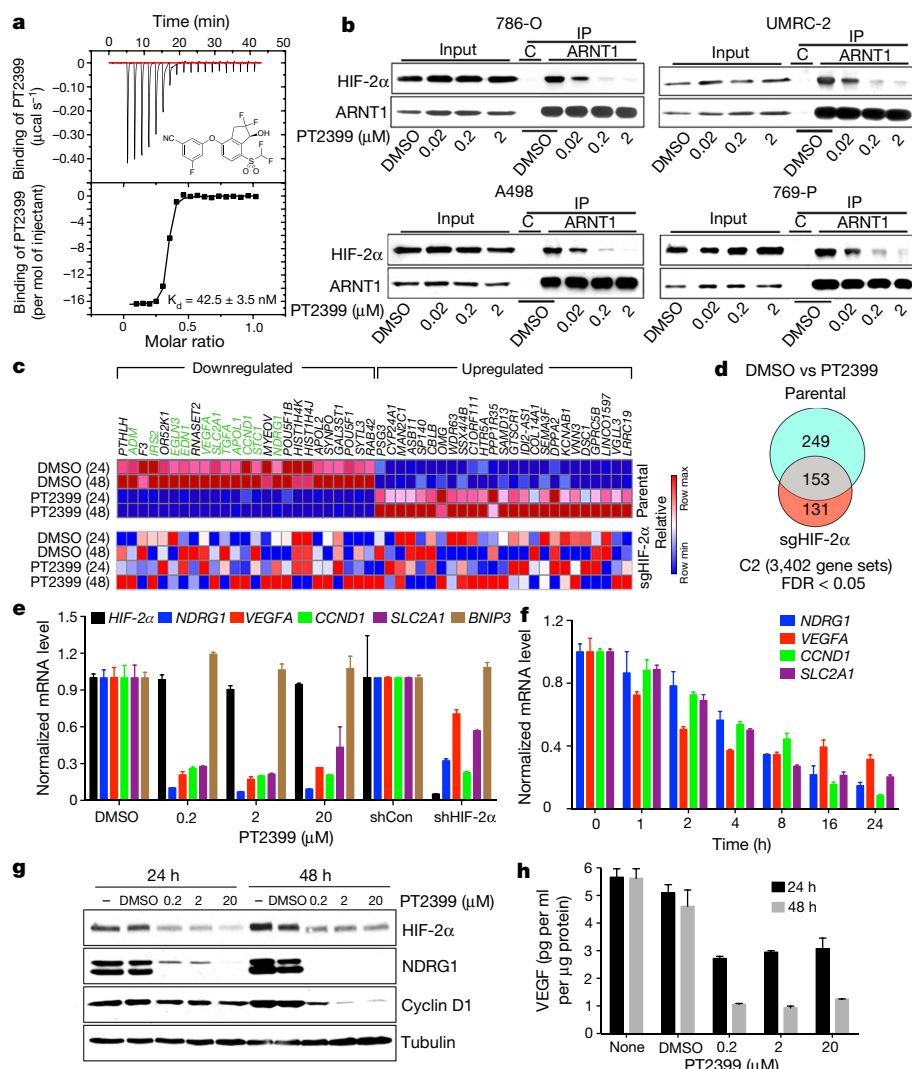


Figure 1 | PT2399 downregulates HIF target genes. **a**, Binding of PT2399 (inset) to human HIF-2 α PAS-B domain assayed by isothermal calorimetry ($n=2$). **b**, Immunoblots of anti-ARNT1 immunoprecipitates (IP) and whole cell extracts (input) from ccRCC cell lines treated with PT2399 or DMSO. C, control IP without ARNT1 antibody. **c**, Heat map of changes in mRNAs in 786-O cells (parental and HIF-2 α single guide RNA (sgHIF-2 α) (guide #6; see Methods)) treated with 2 μ M PT2399 or DMSO for indicated duration (24 or 48 h). $n=1$. Examples of HIF target genes are indicated in green. **d**, Venn diagram of gene sets regulated by PT2399. 249 gene sets were regulated in parental 786-O cells but not in 786-O HIF-2 α cells. **e**, **f**, Levels of the indicated mRNAs, normalized to *ACTB* and then to DMSO controls, in 786-O cells treated with PT2399 at the indicated concentration for 24 h (**e**) or for the indicated duration with 2 μ M PT2399 (**f**). 786-O cells expressing a HIF-2 α shRNA (shHIF-2 α) or control shRNA (shCon) were included in **e** for comparison; $n=3$. **g**, Immunoblots of 786-O cells treated with PT2399. **h**, VEGFA concentration, normalized to total cellular protein, in medium conditioned by 786-O cells treated with PT2399; $n=3$ biological replicates. Data shown as mean \pm s.e.m. (**e**, **f**) or mean \pm s.d. (**h**).

¹Department of Medical Oncology, Dana-Farber Cancer Institute and Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215, USA. ²Peloton Therapeutics, Inc., Dallas, Texas 75235, USA. ³Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235, USA. ⁵Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.

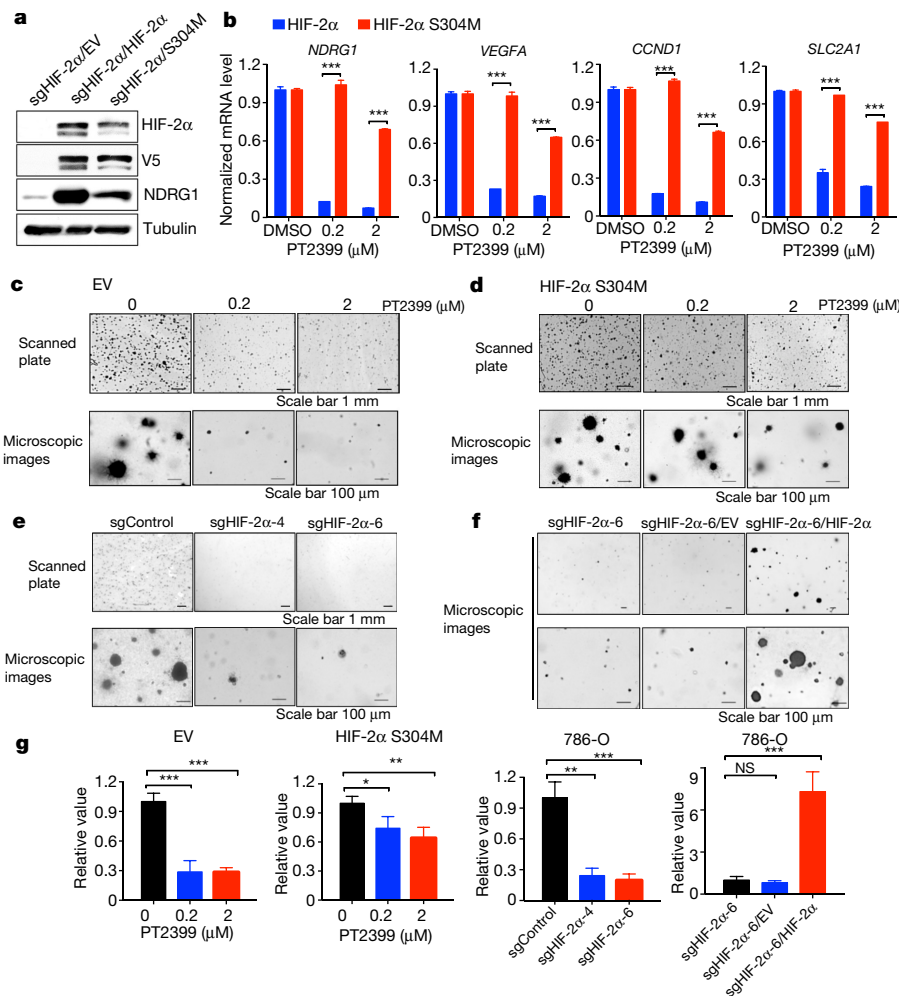


Figure 2 | On-target effects of PT2399 on transcription and growth in soft agar.

a, Immunoblots of HIF-2 α ^{-/-} 786-O cells (sgHIF-2 α #6) infected with a lentivirus encoding V5-tagged HIF-2 α (wild type), HIF-2 α (S304M) or empty vector (EV). **b**, Levels of the indicated mRNAs, normalized to *ACTB* mRNA and then to DMSO treatment, in cells from (a) treated with PT2399 or DMSO for 48 h; $n = 3$. **c, d**, Soft agar colonies formed by 786-O cells infected with a lentivirus encoding the empty vector (c) or HIF-2 α S304M (d) in the presence of PT2399 at the indicated concentrations for 21 days; $n = 3$ biological replicates. **e, f**, Soft agar colonies formed by 786-O clones as in Extended Data Fig. 2a, b; $n = 3$ biological replicates. In **f**, the 786-O HIF-2 α ^{-/-} cells (sgHIF-2 α #6) were superinfected with a lentivirus encoding an sgRNA-resistant HIF-2 α cDNA (HIF-2 α) or with the empty vector (EV). **g**, Quantification of soft agar colony formation in **c–f**. $n = 3$ biological replicates. Data shown as mean \pm s.e.m. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ by two-tailed Student's *t*-tests (b, g). NS, $P > 0.05$.

HIF-2 α from binding to ARNT (Fig. 1b and Extended Data Fig. 1c) and hence to DNA^{2–6}. PT2399 minimally affected a panel of 68 receptors, ion channels, and enzymes (Supplementary Table 4).

Treating *VHL*^{-/-} 786-O clear cell renal cell carcinoma (ccRCC) cells with PT2399 repressed various HIF target genes in mRNA microarray (Fig. 1c), real-time PCR (Fig. 1e, f and Extended Data Fig. 1d), immunoblot (Fig. 1g) and ELISA (Fig. 1h) assays. PT2399 did not suppress HIF-1 α -specific targets such as *BNIP3* (Fig. 1e and Extended Data Fig. 1e). As well as binding directly to HIF-2 α , PT2399 destabilized HIF-2 α , which might enhance the effects of PT2399 on the DNA-binding activity of HIF-2 α (Fig. 1e, g and Extended Data Fig. 1f). PT2399 downregulated gene sets that were induced by hypoxia, HIF, and c-Myc, consistent with reports that HIF-2 α and c-Myc cooperate to promote ccRCC^{7,8} (Fig. 1d, Extended Data Fig. 1g and Supplementary Table 5).

Next we made HIF-2 α ^{-/-} 786-O cells using the CRISPR-Cas9 gene editing technique (Extended Data Fig. 2a). The cells proliferated under standard conditions (Extended Data Fig. 2c–f), consistent with the effects of suppression of HIF-2 α by short hairpin (sh)RNA and treatment with pVHL in 786-O cells^{9,10}. We then used a lentivirus to reintroduce wild-type HIF-2 α or a HIF-2 α missense mutant (S304M) with an occluded PT2399-binding pocket⁴ into these cells (Fig. 2a). The effects of PT2399 on HIF-responsive mRNAs was largely eliminated in cells lacking HIF-2 α (Fig. 1c, d) or expressing HIF-2 α S304M (Fig. 2b).

PT2399 (up to 2 μ M) minimally altered ccRCC cell line proliferation under standard cell culture conditions (Extended Data Fig. 2c, g–j). A larger dose of PT2399 (20 μ M) caused off-target toxicity because it inhibited the proliferation of HIF-2 α ^{-/-} 786-O cells (Extended Data

Fig. 2a, d–f) and other cancer cell lines with undetectable HIF-2 α (Extended Data Fig. 2k–m). However, PT2399 inhibited 786-O cell growth in soft agar at doses of 0.2–2 μ M (Fig. 2c, g and Extended Data Fig. 3a, f). This effect was specific because it was reversed by expression of HIF-2 α S304M (Fig. 2d, g) and not seen in *VHL*^{+/+} SLR21 ccRCC cells (Extended Data Fig. 3c, f). Similarly, HIF-2 α ^{-/-} 786-O cells did not form soft agar colonies unless rescued with exogenous HIF-2 α (Fig. 2e–g and Extended Data Fig. 2b). Therefore, PT2399 decreases HIF-dependent transcription and soft agar growth in an on-target manner.

As a step towards imaging studies we infected 786-O cells, as well as isogenic cells expressing exogenous pVHL, with a lentivirus encoding firefly luciferase (Luc) driven by a HIF-responsive promoter (3 \times HRE-Luc). As expected, PT2399 inhibited Luc activity in the *VHL*^{-/-} cells but not in their pVHL-proficient counterparts (Fig. 3a, b). Conversely, the dioxygenase inhibitor dimethylxaloylglycine (DMOG), which blocks the binding of pVHL to HIF α , induced Luc activity in the pVHL-proficient cells but not in the *VHL*^{-/-} cells (Fig. 3a). As expected, PT2399 did not affect Luc driven by constitutive promoters, such as the CMV promoter (Fig. 3c).

Next, 3 \times HRE-Luc 786-O cells and CMV-Luc 786-O cells were injected into opposing kidneys of nude mice. Once tumours were established, as determined by serial bioluminescence imaging (BLI), the mice were given PT2399 or vehicle twice daily (Extended Data Fig. 1h, i). Two days of treatment with PT2399 decreased the 3 \times HRE-Luc signal by more than 60%, similar to its effects *in vitro* (Fig. 3d, f). These effects were not observed in the CMV-Luc tumours or in vehicle-treated mice (Fig. 3d, e). In PT2399-treated mice, the 3 \times HRE-Luc signal recovered after a drug washout period and

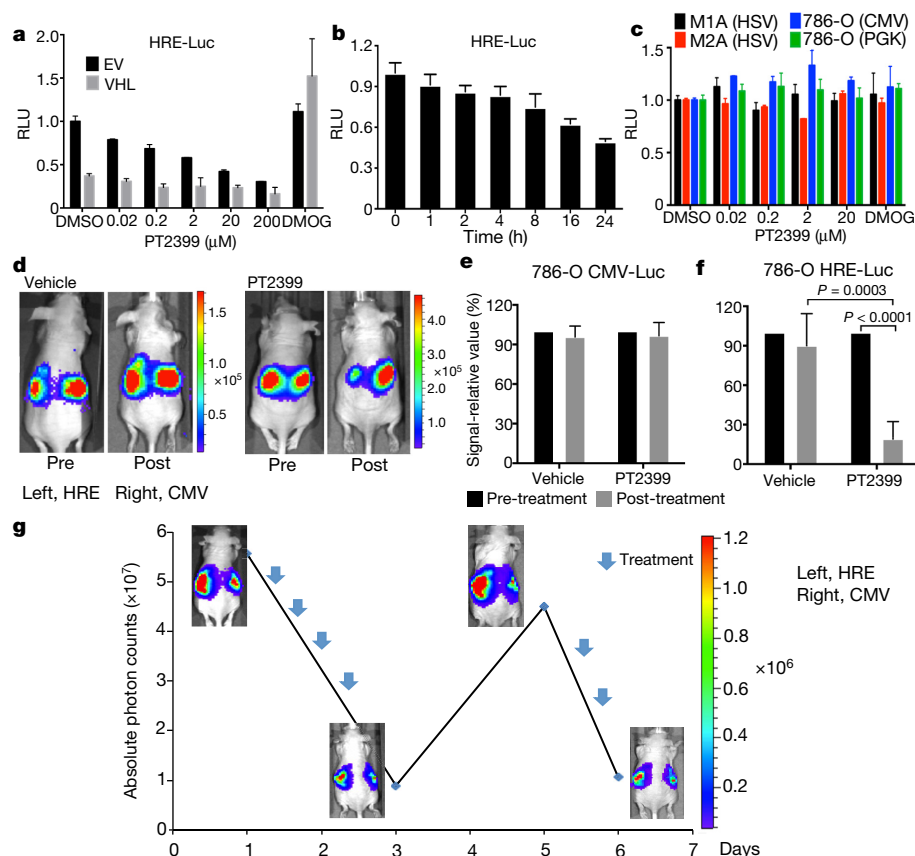


Figure 3 | Pharmacodynamic effects of PT2399 *in vivo*. **a, b**, Light emission, normalized to total cellular protein (**a**) and cell number (**b**) in 786-O 3 × HRE-Luc reporter cells expressing VHL or an empty vector (EV). The cells were treated with PT2399 overnight (**a**) or at 2 μM for the indicated durations (**b**). DMOG (1 mM) was included in **a** as a control; *n* = 3. RLU, relative luciferase units. **c**, RLU values of 786-O derivatives expressing luciferase driven by the indicated promoters and then exposed to either PT2399 or 1 mM DMOG; *n* = 3 biological replicates. **d**, Representative bioluminescent images (BLI) of mice with orthotopic tumours formed by 786-O 3 × HRE-Luc reporter cells (left kidney) or 786-O CMV-Luc reporter cells (right kidney). Images were obtained before and after two days of treatment with 30 mg kg⁻¹ PT2399 given twice daily (*n* = 6 mice from two independent experiments) or vehicle (*n* = 6 mice from two independent experiments) by oral gavage. **e, f**, Quantification of BLI from mice as in **d**; *n* = 4 mice from two independent experiments. Values were normalized to the pretreatment values for each treatment. **g**, Serial BLI of a mouse as in **d** treated with PT2399 (blue arrows). Shown on the y-axis are absolute photon counts for the 3 × HRE-Luc tumour. Data shown as mean ± s.e.m. (**a–c, e, f**). Statistical significance was assessed using Mann–Whitney tests (**f**).

decreased again after drug rechallenge (Fig. 3g). Analysis of kidneys removed after 2 days of PT2399 treatment revealed decreased HIF-responsive mRNAs, decreased Ki-67 staining (representing decreased proliferation), increased caspase 3 cleavage (reflective of increased apoptosis) and decreased microvessel density (Extended Data Fig. 4a–e).

To assess the antitumour efficacy of PT2399, CMV-Luc 786-O cells were grown orthotopically in nude mice; once tumours were established, the mice were treated with PT2399 or vehicle. As expected, tumours continued to grow in vehicle-treated mice, as shown by weekly BLI. In contrast, PT2399 caused tumour stasis or regression (Fig. 4a–c and Extended Data Fig. 4f), which correlated with decreased circulating tumour-derived VEGF (Extended Data Fig. 4g), decreased proliferation and decreased angiogenesis (Extended Data Fig. 4h).

Kidney-confined ccRCC can often be treated surgically. We therefore obtained a metastatic variant of 786-O cells (M2A cells) expressing Luc under an HSV TK promoter; these cells form diffuse lung colonies after tail vein injection¹¹. In this model, PT2399 still caused marked tumour regression and prolonged survival (Fig. 4d–f). Similar results were obtained when a limited number of cells were injected in an effort to better mimic established lung metastases (Extended Data Fig. 5a, b). Introducing HIF-2α S304M into M2A cells (Fig. 4g) conferred partial resistance to the pharmacodynamic (Extended Data Fig. 5c) and antitumour effects of PT2399 (Fig. 4h, i).

PT2399 also inhibited the growth of *VHL*^{-/-} A498 ccRCC cells in soft agar and orthotopic tumour assays (Extended Data Fig. 3b, f and Extended Data Fig. 6), consistent with the effects of HIF-2α shRNA in these cells¹⁰, and inhibited the growth of a *VHL*^{-/-} ccRCC patient-derived xenograft (PDX) (Extended Data Fig. 5d, e). By contrast, PT2399 did not suppress orthotopic tumours formed by *VHL*^{-/-} UMRC-2 cells or *VHL*^{-/-} 769-P ccRCC cells (Fig. 5a–d) despite inhibiting HIF-2α dimerization and HIF-2α-dependent transcription in these cells with IC₅₀s comparable to those seen in 786-O and A498 cells (Fig. 1b and Extended Data Fig. 7a–d) and

despite effective suppression of HIF target genes *in vivo* (Extended Data Fig. 8a).

To study this differential sensitivity further, we measured HIF-2α abundance and the response of selected HIF target genes to PT2399 and to pVHL across a panel of *VHL*^{-/-} ccRCC cell lines. The PT2399-sensitive 786-O and A498 cells had higher HIF-2α levels than the insensitive UMRC-2 and 769-P cells (Fig. 5e). The PT2399-sensitive cells also exhibited greater inhibition of HIF target genes, as a percentage of basal expression, in response to PT2399 (Fig. 5f and Extended Data Fig. 8d, e) and, where tested, pVHL (Extended Data Fig. 8b, c). The latter finding further supported the idea that this differential sensitivity to PT2399 reflects differences in HIF-2α dependence rather than differences in intracellular drug accumulation. Indeed, growth of UMRC-2 cells and 769-P cells in soft agar was insensitive to Cas9-mediated inactivation of HIF-2α and to PT2399, unlike growth of 786-O and A498 cells (Fig. 5g–j and Extended Data Fig. 3c, f, g). Similarly, growth of *VHL*^{-/-} SKRC-20 and UMRC-6 ccRCC cells in soft agar was unaffected by genetic disruption of HIF-2α or treatment with PT2399 (Extended Data Fig. 3c–g). Growth of RCC10 cells in soft agar was unaffected by PT2399, but was suppressed in an on-target manner by Cas9-mediated loss of HIF-2α, despite the cells showing a similar decrease in HIF-2 target mRNAs in response to both (Extended Data Fig. 3c–g and Extended Data Fig. 9a–f). The importance of this discrepancy is unclear because the *HIF-2α*^{-/-} RCC10 cells quickly regained the ability to grow in soft agar after repeated culture despite persistent HIF-2α loss (see also below). Finally, we confirmed that genetic disruption of HIF-2α, like treatment with PT2399, did not affect orthotopic tumour growth by UMRC-2 cells (Extended Data Fig. 10).

Differential HIF-2α dependence amongst ccRCC lines is not linked to their HIF-1α status because the insensitive cell lines 769-P and SKRC-20, like the sensitive cell lines 786-O and A498, lack wild-type HIF-1α¹². Moreover, Cas9-mediated ablation of HIF-1α in UMRC-2 cells did not render them HIF-2α-dependent in soft agar assays (Extended Data Fig. 7e–g).

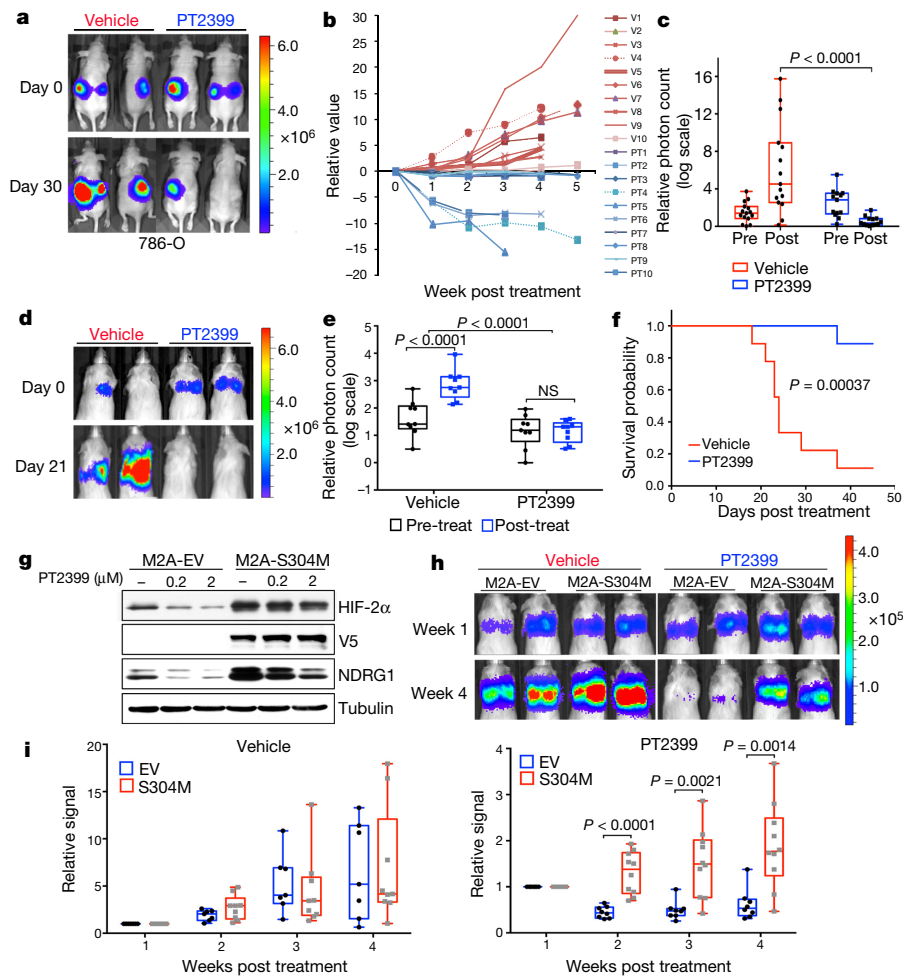


Figure 4 | PT2399 antitumour activity. **a**, Representative BLI of orthotopic tumours formed by CMV-Luc 786-O cells before and after treatment with PT2399 (30 mg kg⁻¹) or vehicle twice daily by oral gavage for 30 days (vehicle, $n = 15$ mice from three independent experiments; PT2399, $n = 13$ mice from three independent experiments). **b**, Spider plot showing growth of tumours as in **a** determined by serial BLI. For each tumour the BLI values were normalized to the corresponding day 0 value. V, tumours from vehicle-treated mice; PT, tumours from PT2399-treated mice. **c**, Quantification of BLI from mice as in **a**. The value for each tumour was normalized to the pretreatment value for that tumour. **d**, Representative BLI of lung colonies formed by M2A-Luc cells in mice treated with PT2399 (30 mg kg⁻¹) or vehicle twice daily by oral gavage (vehicle, $n = 9$; PT2399, $n = 9$ from two independent experiments).

By contrast, HIF-2 α dependence across the ccRCC lines we examined loosely correlated with their basal HIF-2 α levels and the dependence of HIF target genes in those lines on HIF-2 α itself. A caveat is that some of these lines might have lost their dependence on HIF-2 α owing to prolonged passage in culture, especially as HIF-2 α is not required under standard culture conditions. On the other hand, freshly explanted ccRCC PDXs also show variable sensitivity to PT2399 (ref. 13).

To begin to understand this differential HIF-2 α dependence further, we focused on RCC10 cells because HIF target genes and soft agar growth are less dependent on HIF-2 α in these cells than in 786-O and A498 cells despite their comparably high HIF-2 α levels (Fig. 5e and Extended Data Figs 3c–g, 9b). We discovered that RCC10 cells harbour the canonical p53 R248W mutant (Extended Data Fig. 9g). The p53 R248W mutation also arose spontaneously in a 786-O subclone made in our laboratory (Extended Data Fig. 9h) and was associated with acquired resistance to PT2399 (Extended Data Fig. 9i). p53 was not induced by DNA damage in the HIF-2-independent lines UMRC-2

e, **f**, Quantification of BLI values (**e**) and Kaplan–Meier survival curves (**f**) from mice treated as in **d** (vehicle, $n = 9$ mice from two independent experiments; PT2399, $n = 9$ mice from two independent experiments). **g**, Immunoblots of M2A cells infected with a lentivirus encoding HIF-2 α S304M or empty vector (EV). **h**, Representative BLI of lung colonies formed by M2A cells in **g** treated with PT2399 (30 mg kg⁻¹) or vehicle twice daily by oral gavage (for empty vector, $n = 7$ vehicle-treated and $n = 8$ PT2399-treated; for S304M, $n = 9$ vehicle-treated and $n = 10$ PT2399-treated) from two independent experiments. **i**, Quantification of BLI values as in **h**. Data shown as median with range (**c**, **e**, **i**). Statistical significance was assessed using Mann–Whitney test (**c** and **i**), unpaired t -test (**e**) or log-rank test (**f**).

and 769-P, suggesting that these cells also have p53 pathway mutations, but was induced in the HIF-2-independent lines UMRC-6 and Caki-2 (Extended Data Fig. 9g, j). Of note, p53 was modestly induced by PT2399 and Cas9-mediated loss of HIF-2 α in p53^{+/+} 786-O cells (Extended Data Fig. 9h, k), consistent with reports that HIF-2 α constrains p53 activity in ccRCC^{14,15}. Therefore an intact p53 pathway seems necessary, but not sufficient, for HIF-2-dependence in ccRCC.

Our findings suggest that the response of *VHL*^{-/-} ccRCC to HIF-2 α antagonists will be variable and that predictive biomarkers, perhaps including measures of HIF-2 α activity and p53 status, will be needed. The current view that p53 mutations are uncommon in ccRCC is based mainly on studies of primary tumours removed at nephrectomy¹. p53 pathway mutations might be more common in metastatic ccRCC, from which most ccRCC lines are derived, or might arise after ccRCC therapies¹⁶. Alternatively, some p53 mutations in ccRCC lines might have arisen *ex vivo*. Another important question is whether adding HIF-2 α antagonists will enhance the activity of existing ccRCC drugs.

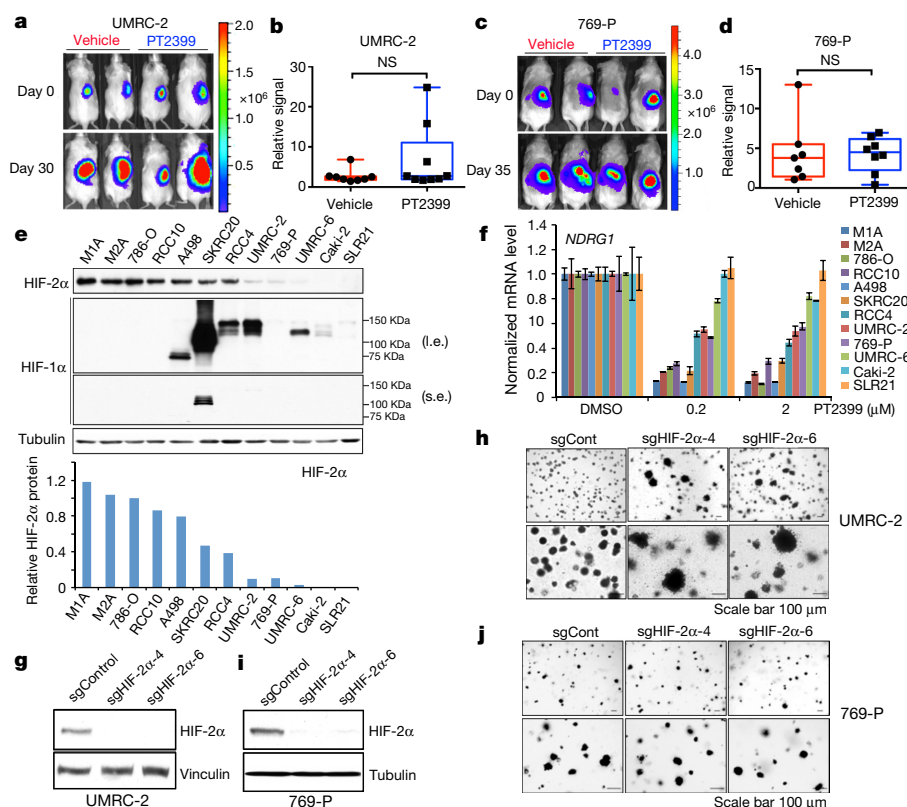


Figure 5 | Variable sensitivity of ccRCC lines to PT2399. **a, c,** Representative BLI of orthotopic tumours formed by CMV-Luc UMRC-2 cells (**a**) (vehicle, $n = 8$ and PT2399, $n = 9$ mice from two independent experiments) and CMV-Luc 769-P cells (**c**) (vehicle, $n = 7$ and PT2399, $n = 8$ mice from two independent experiments) before and after (30 and 35 days, respectively) treatment with PT2399 (30 mg kg⁻¹) or vehicle by oral gavage. **b, d,** Quantification of BLI as in **a** and **c**, respectively. **e,** Immunoblots of ccRCC cell lines, including quantification of HIF-2 α protein levels normalized to tubulin. HIF-1 α in A498 and SKRC20 cells is mutated and defective¹⁴. i.e., long exposure; s.e., short exposure. **f,** *NDRG1* mRNA levels in ccRCC cell lines treated with PT2399 for 24 h; $n = 3$ biological replicates. For each cell line, after normalization to *ACTB*, *NDRG1* mRNA levels were normalized to the untreated value for that cell line. *VHL*^{+/+} SLR21 renal carcinoma cells were included in **e** and **f** for comparison. **g, i,** Immunoblots of UMRC-2 cells (**g**) and 769-P cells (**i**) after CRISPR-based gene editing with control sgRNA or HIF-2 α sgRNA (guides 4 and 6). **h, j,** Soft agar colonies formed by cells as in **g** and **i**; $n = 3$ biological replicates. Data shown as median with range (**b, d**) and mean \pm s.d. (**f**). Statistical analysis was performed using two-tailed Student's *t*-tests (**b, d**). NS, $P > 0.05$.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 September 2015; accepted 22 August 2016.

Published online 5 September 2016.

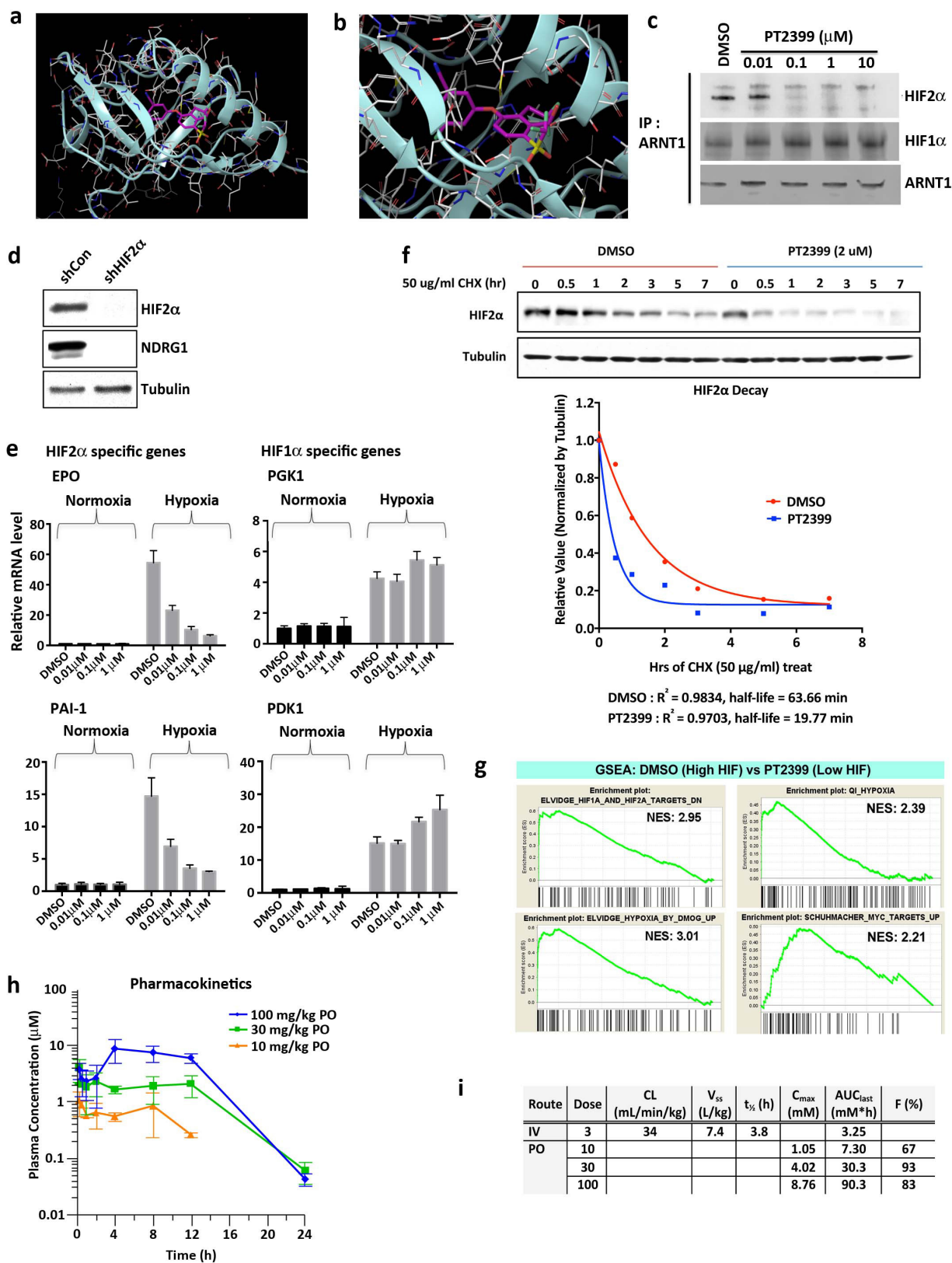
- Kaelin, W. in *Kidney Cancer: Principles and Practice* Ch. 3, 31–57 (Springer International Publishing, 2015).
- Scheuermann, T. H. et al. Artificial ligand binding within the HIF2 α PAS-B domain of the HIF2 transcription factor. *Proc. Natl Acad. Sci. USA* **106**, 450–455 (2009).
- Scheuermann, T. H. et al. Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat. Chem. Biol.* **9**, 271–276 (2013).
- Rogers, J. L. et al. Development of inhibitors of the PAS-B domain of the HIF-2 α transcription factor. *J. Med. Chem.* **56**, 1739–1747 (2013).
- Dixon, D. D. et al. Aryl ethers and uses thereof. US Patent WO2015035223 (2015).
- Wallace, E. M. et al. A small-molecule antagonist of HIF2 α is efficacious in preclinical models of renal cell carcinoma. *Cancer Res.* **15**, 5491–5500 (2016).
- Gordan, J. D. et al. HIF- α effects on c-Myc distinguish two subtypes of sporadic VHL-deficient clear cell renal carcinoma. *Cancer Cell* **14**, 435–446 (2008).
- Gordan, J. D., Bertout, J. A., Hu, C. J., Diehl, J. A. & Simon, M. C. HIF-2 α promotes hypoxic cell proliferation by enhancing c-myc transcriptional activity. *Cancer Cell* **11**, 335–347 (2007).
- Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G., Jr. Tumour suppression by the human von Hippel-Lindau gene product. *Nat. Med.* **1**, 822–826 (1995).
- Kondo, K., Kim, W. Y., Lechpammer, M. & Kaelin, W. G. Jr. Inhibition of HIF2 α is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol.* **1**, E83 (2003).
- Vanharanta, S. et al. Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nat. Med.* **19**, 50–56 (2013).
- Shen, C. et al. Genetic and functional studies implicate HIF1 α as a 14q kidney cancer suppressor gene. *Cancer Discov.* **1**, 222–235 (2011).
- Chen, W. et al. Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* <http://dx.doi.org/10.1038/nature19796> (2016).
- Bertout, J. A. et al. HIF2 α inhibition promotes p53 pathway activity, tumor cell death, and radiation responses. *Proc. Natl Acad. Sci. USA* **106**, 14391–14396 (2009).
- Roberts, A. M. et al. Suppression of hypoxia-inducible factor 2 α restores p53 activity via Hdm2 and reverses chemoresistance of renal carcinoma cells. *Cancer Res.* **69**, 9056–9064 (2009).
- Fay, A. P. et al. Whole-exome sequencing in two extreme phenotypes of response to VEGF-targeted therapies in patients with metastatic clear-cell renal cell carcinoma. *J. Natl Comp. Cancer Network* **14**, 820–824 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Massague for 786-M1A and M2A cells. Supported by grants from the NIH. W.G.K. is an HHMI Investigator. R.K.B. is the Michael L. Rosenberg Scholar in Medical Research and was supported by the Cancer Prevention and Research Institute of Texas (RP130513).

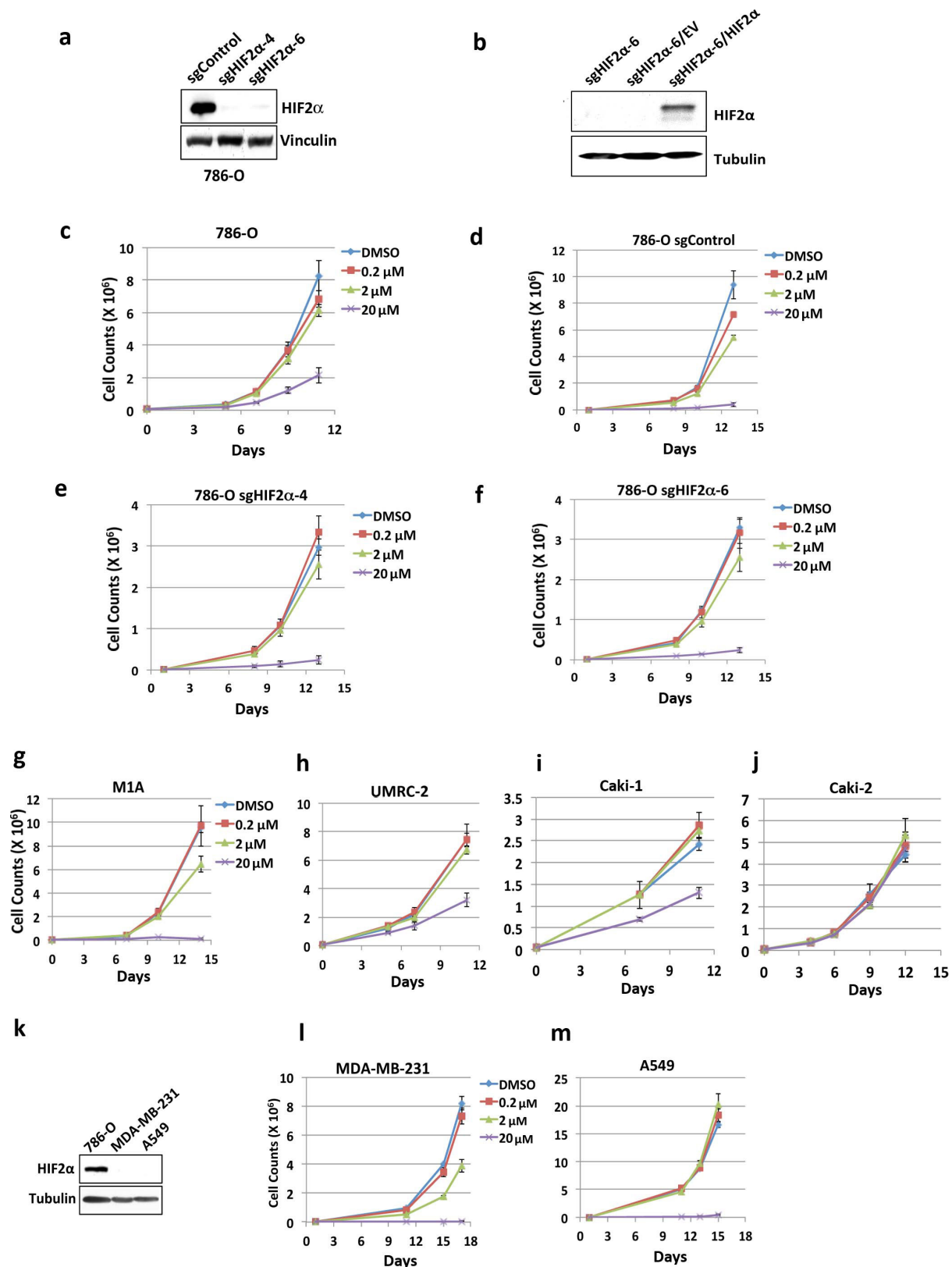
Author Contributions W.G.K. conceived the study, analysed data and, with H.C., wrote the manuscript; H.C. designed and did all of the experiments, except for the structural, isothermal titration calorimetry, and pharmacokinetic studies and the experiments for Extended Data Figs 1c, e; X.D. generated proteins, performed ITC experiments and collected and solved the HIF-2–PT2399 co-crystal structure; J.P.R., J.A.J. and E.M.W. coordinated discovery, synthesis and characterization of PT2399; J.A.J. and E.M.W. designed the PDX experiment; E.L. assisted with lung colonization assays; A.A.C. carried out bioinformatic analyses; W.G. designed and generated CRISPR reagents; I.C. and S.S. performed and analysed immunohistochemistry; R.B. designed the PT2399-resistant HIF-2 α .

Author Information The crystal structure of PT2399 bound to the HIF-2 α -B*–ARNT-B* complex has been deposited in the Protein Data Bank under accession number 5TOT. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.G.K. (william_kaelin@dfci.harvard.edu) or E.M.W. (eli.wallace@pelotontx.com).



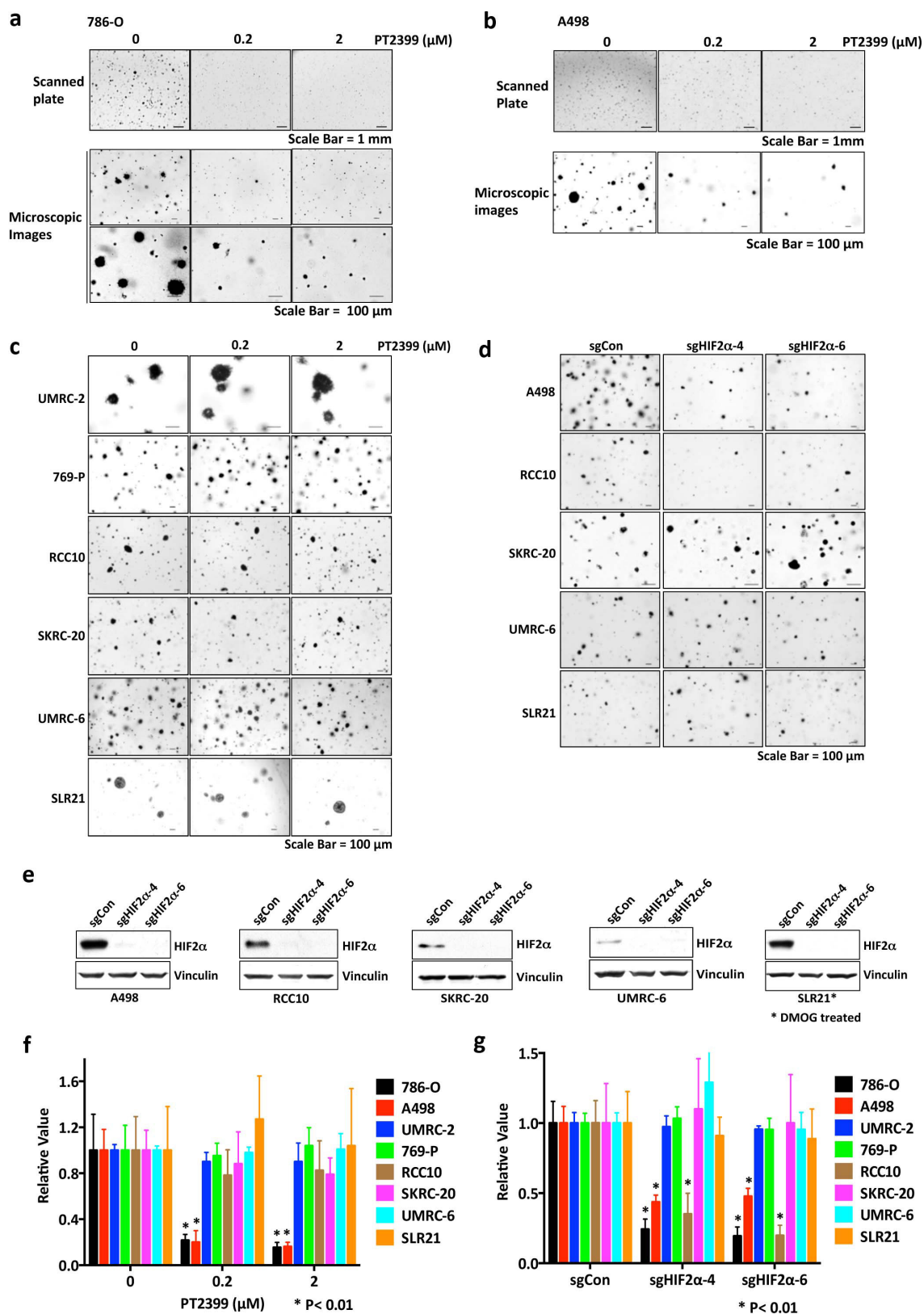
Extended Data Figure 1 | Binding of PT2399 to PAS-B domain of human HIF-2α as determined by X-ray co-crystal structure. **a**, X-ray co-crystal of PT2399 (magenta) bound to complex formed by HIF-2α and ARNT PAS-B domains (ARNT removed for clarity). **b**, X-ray co-crystal of PT2399 (magenta) with HIF-2α and ARNT PAS-B domains (zoomed in on HIF-2α PAS-B pocket). **c**, Immunoblots of anti-ARNT1 immunoprecipitates (IP) of Hep3B cells treated with PT2399 or DMSO. **d**, Immunoblot of 786-O cells expressing shRNA against HIF-2α (3806)

or control shRNA. **e**, HIF-2α-specific gene regulation in Hep3B cells; $n = 3$ biological replicates. **f**, Immunoblot analysis (top) and quantification (bottom) of HIF-2α in 786-O cells treated with DMSO or PT2399 for 16 h and then exposed to cycloheximide for the indicated time periods; $n = 3$ biological replicates. **g**, Enrichment plots for representative gene sets previously linked to HIF, hypoxia, or c-Myc. **h**, **i**, Plasma PT2399 levels after administration of a single dose of PT2399 to CD-1 mice; $n = 3$ per time point from one experiment.



Extended Data Figure 2 | Inhibition of cell proliferation by PT2399 *ex vivo*. **a, b**, Immunoblot analysis of 786-O cells after CRISPR-based gene editing with control sgRNA or HIF-2α sgRNA (guides 4 and 6). In **b**, cells were also infected with an empty vector (EV) or a virus expressing an HIF-2α sgRNA guide 6-resistant HIF-2α cDNA. **c–f**, Cell proliferation of parental 786-O cells (**c**) and 786-O clones subjected to CRISPR-based gene editing with a control sgRNA (**d**) or in which endogenous HIF-2α was

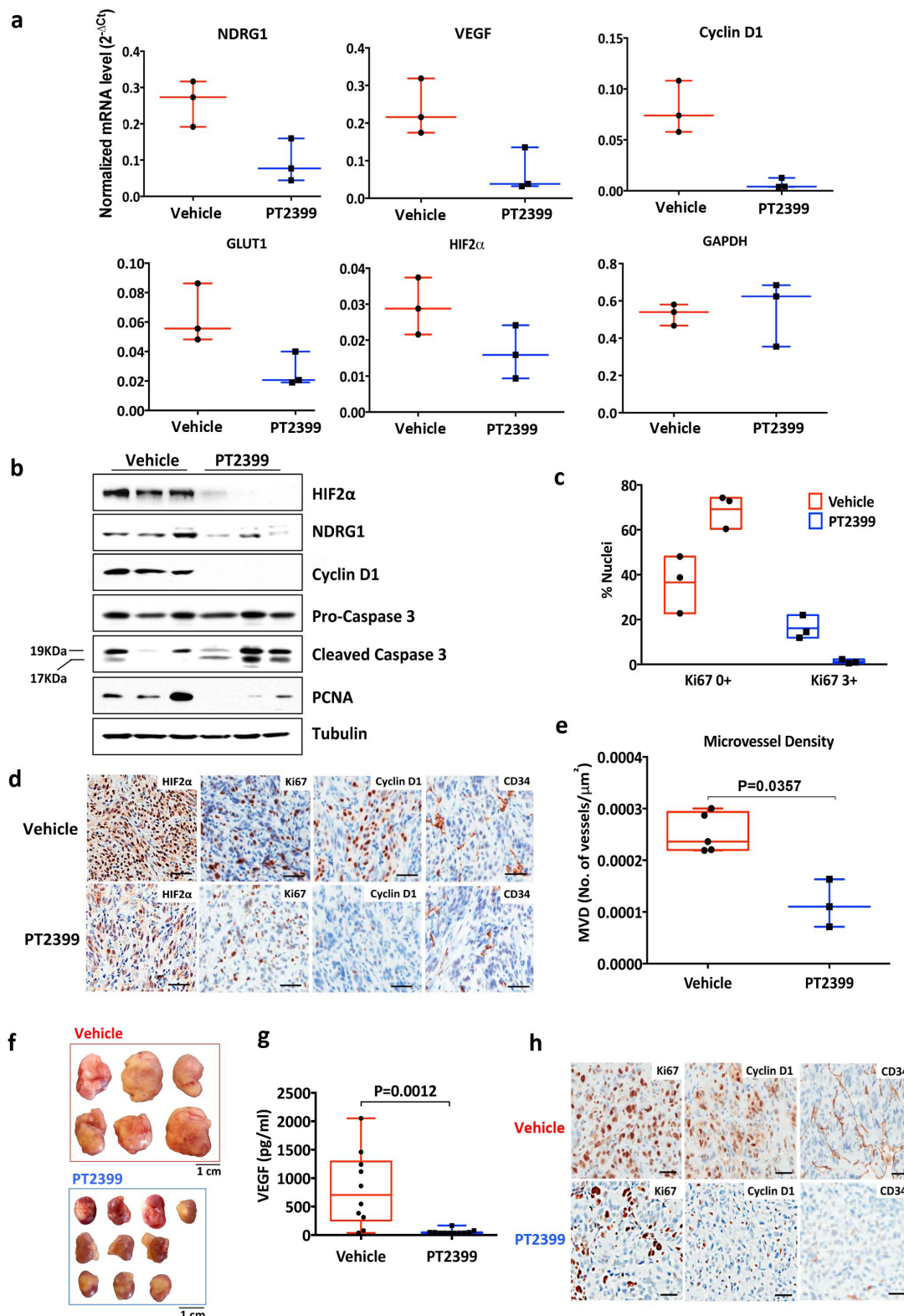
successfully inactivated using two different HIF-2α sgRNAs (guides 4 and 6) (**e, f**); $n = 3$ biological replicates. **g–j**, Proliferation curves for 786-M1A (**g**), UMRC-2 (**h**), Caki-1 (**i**), and Caki-2 cells (**j**) treated with the indicated concentrations of PT2399. **k**, Immunoblot analysis of the indicated cell lines. **l, m**, Proliferation curves for MDA-MB-231 (**l**) and A549 cells (**m**) treated with the indicated concentrations of PT2399; $n = 3$ biological replicates. Data shown as mean \pm s.d. (**c–j, l, m**).



Extended Data Figure 3 | Effects of PT2399 on soft agar growth.

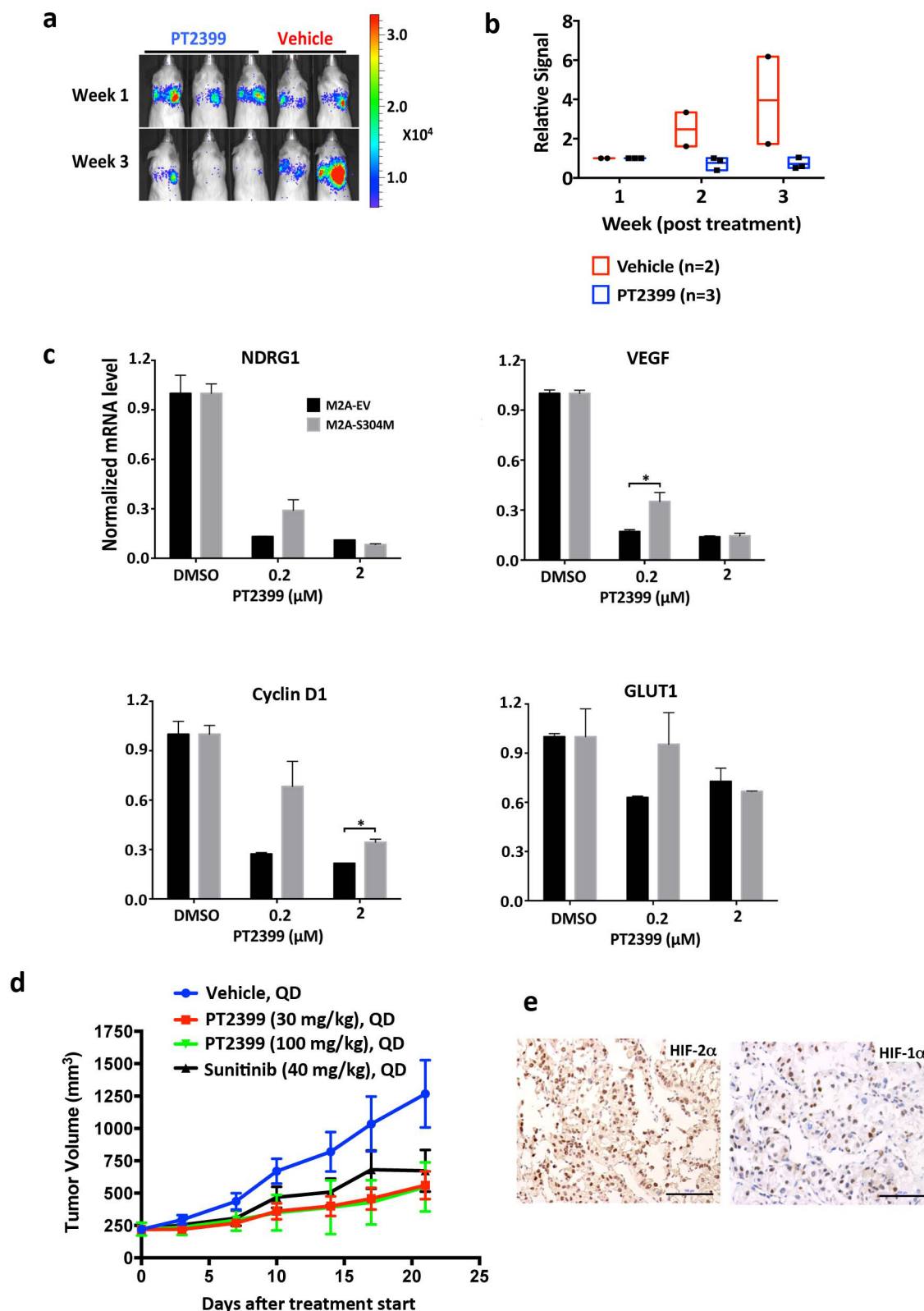
a–c, Soft agar colonies formed by 786-O cells (**a**), A498 cells (**b**), and the indicated cell lines (**c**) in the presence of PT2399 at the indicated concentrations for 21 days; $n = 3$ biological replicates. **d**, Soft agar colonies formed by the indicated polyclonal cell line populations after CRISPR-based gene editing with control sgRNA or HIF-2 α sgRNA (guides 4 and 6);

$n = 3$ biological replicates. The reason for the differential sensitivity of RCC10 cells to PT2399 and the HIF-2 α sgRNAs is not yet clear. **e**, Immunoblot analysis of the cells used in **d**. For SLR21 cells, 1 mM DMOG was added for 16 h to detect HIF-2 α . **f**, **g**, Quantification of soft agar colonies formed in **a–d** and Fig. 5h, j, respectively; $n = 3$. Data shown as mean \pm s.e.m. * $P < 0.01$ by two-tailed Student's t -tests (**f**, **g**).



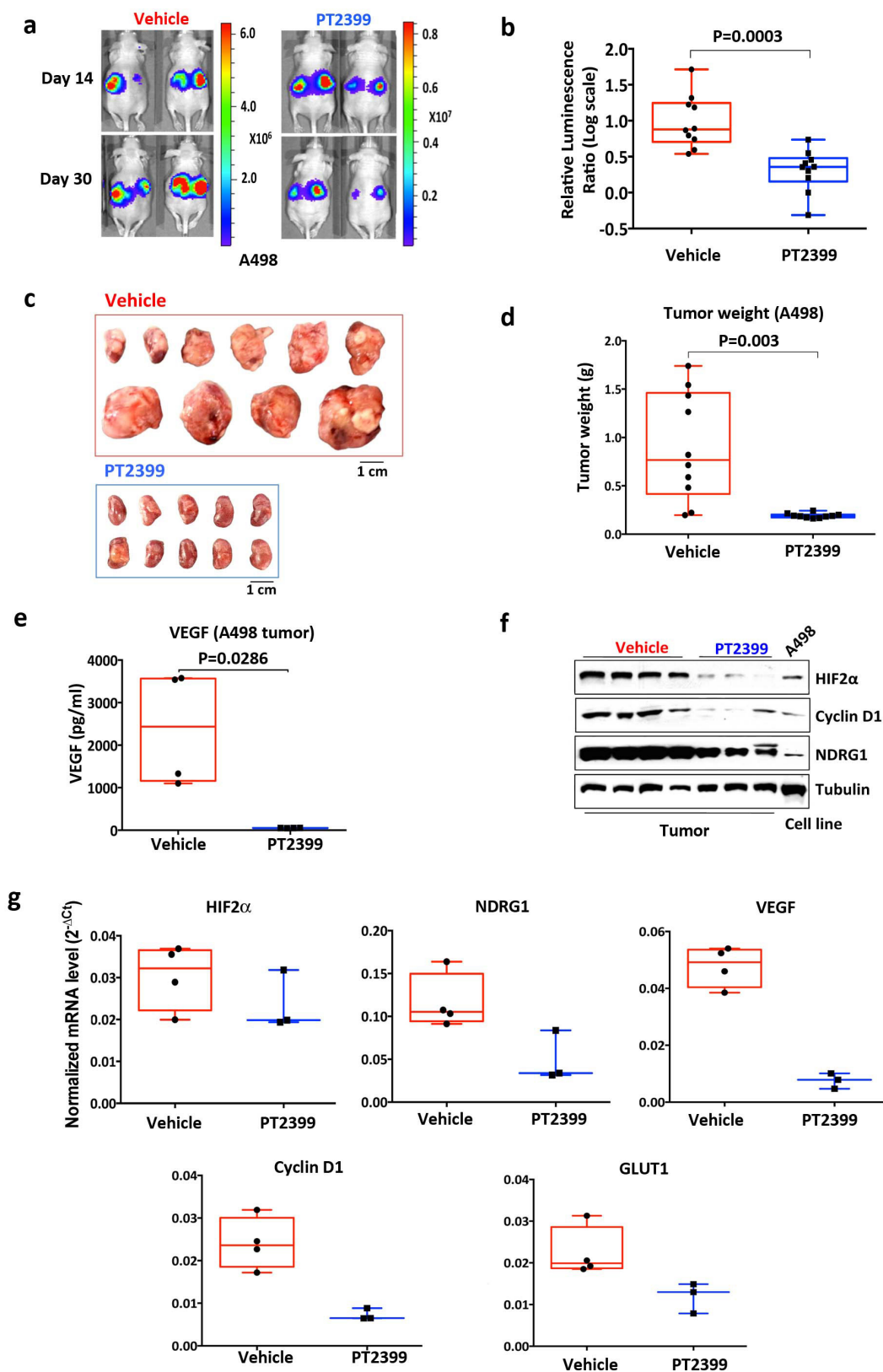
Extended Data Figure 4 | Pharmacodynamic effects of PT2399 *in vivo*. **a**, Levels of the indicated mRNAs, normalized to *ACTB*, in 786-O orthotopic tumours treated with PT2399 (30 mg kg^{-1}) ($n = 3$ mice from two independent experiments) or vehicle ($n = 3$ mice from two independent experiments) twice daily for two days *in vivo*. **b**, Immunoblot analysis of 786-O orthotopic tumours treated with PT2399 (30 mg kg^{-1}) or vehicle twice daily for two days *in vivo*; vehicle, $n = 3$ mice from two independent experiments; PT2399, $n = 3$ mice from two independent experiments. **c**, Quantification of Ki-67 staining (vehicle, $n = 3$; PT2399, $n = 3$ mice from two independent experiments). **d**, Immunohistochemistry of representative 786-O orthotopic tumours treated with PT2399

(30 mg kg^{-1}) or vehicle twice daily for two days *in vivo*; for vehicle, $n = 3$ and PT2399, $n = 3$. Scale bars, $50 \mu\text{m}$. **e**, Microvessel density (vehicle, $n = 5$; PT2399, $n = 3$ mice from two independent experiments) from tumours as in **d**. **f**, **g**, Representative tumours at necropsy (**f**) and serum VEGF concentrations (vehicle, $n = 10$; PT2399, $n = 11$ mice from three independent experiments) (**g**) from mice as in Fig. 4a–c just before necropsy. **h**, Representative immunohistochemical staining of 786-O tumours treated as in Fig. 4a–c (vehicle, $n = 4$; PT2399, $n = 5$); Scale bars, $50 \mu\text{m}$. Data shown as median with range (**a**, **c**, **e**, **g**). Statistical significance was assessed by using Mann–Whitney test (**e**) or unpaired *t*-test (**g**). NS, $P > 0.05$.



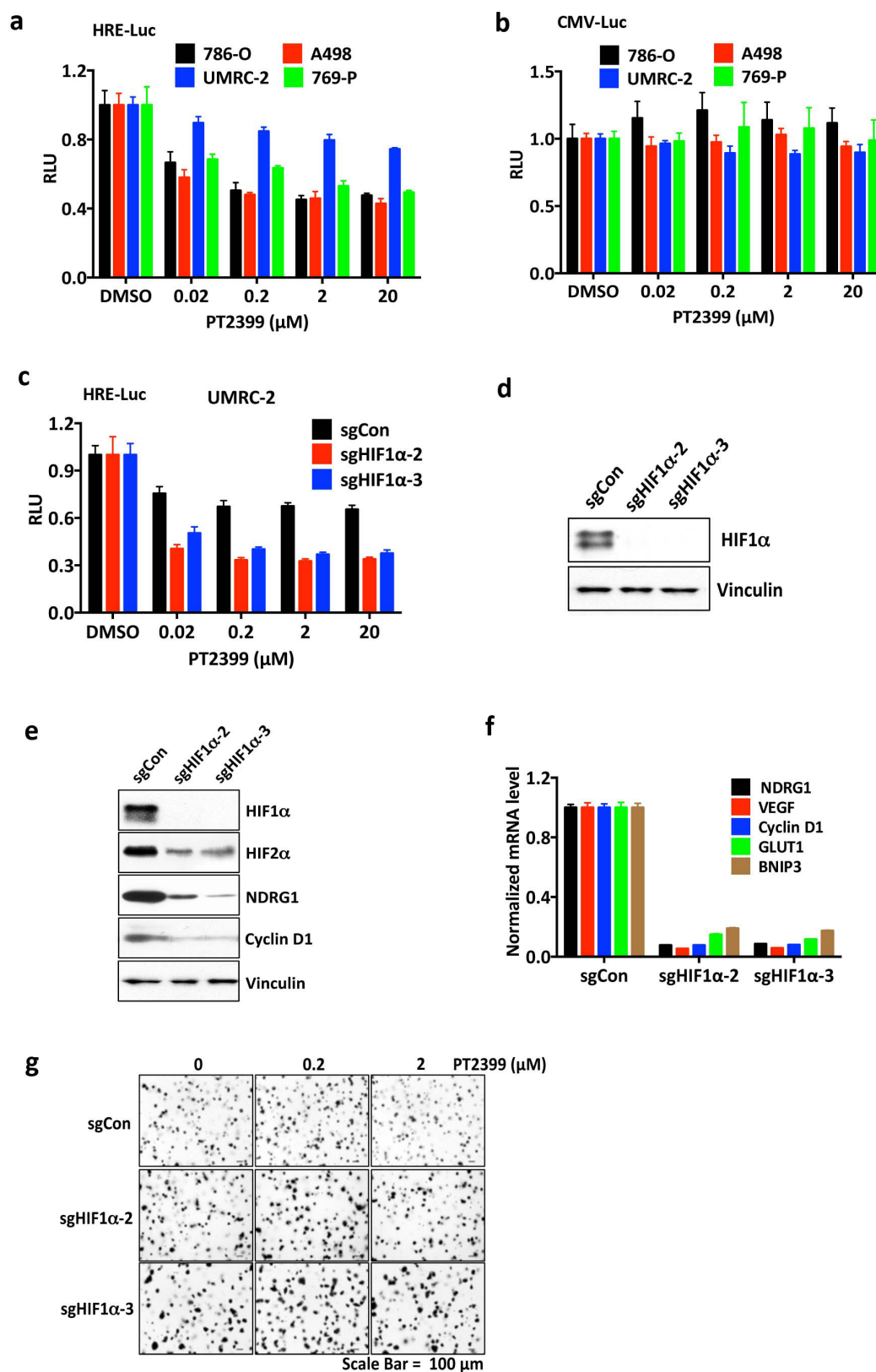
Extended Data Figure 5 | Antitumour activity of PT2399 in lung colonization and PDX models. **a**, BLI of lung colonies formed after tail vein injection of 9,000 786-M2A cells treated with PT2399 (30 mg kg⁻¹) or vehicle twice daily by oral gavage. Treatment began at week 1. **b**, Quantification of BLI values as in **a**. Data shown as median with range (vehicle, $n = 2$ and PT2399, $n = 3$ mice from one experiment). **c**, Partial rescue of PT2399 pharmacodynamic effect by HIF-2 α S304M; $n = 3$ biological replicates. Levels of the indicated mRNAs, normalized to *ACTB* mRNA and then to DMSO treatment, in cells from Fig. 4g treated with

PT2399 at the indicated concentrations for 48 h; $n = 3$. Data shown as mean \pm s.e.m. * $P < 0.05$ by two-tailed Student's *t*-tests. Note that rescue is only partial, perhaps because these cells still produce endogenous wild-type HIF-2 α in addition to exogenous HIF-2 α S304M. **d**, Subcutaneous PDX measurements in mice randomized to the indicated treatments, including the FDA-approved ccRCC drug sunitinib, when the tumours reached 200–300 mm³. $P < 0.05$ for difference between PT2399 and vehicle ($n = 8$, unpaired *t*-test). **e**, Immunohistochemistry of PDX in **d** before treatment. Scale bars, 100 μ m. Data shown as mean \pm s.e.m. (**d**).



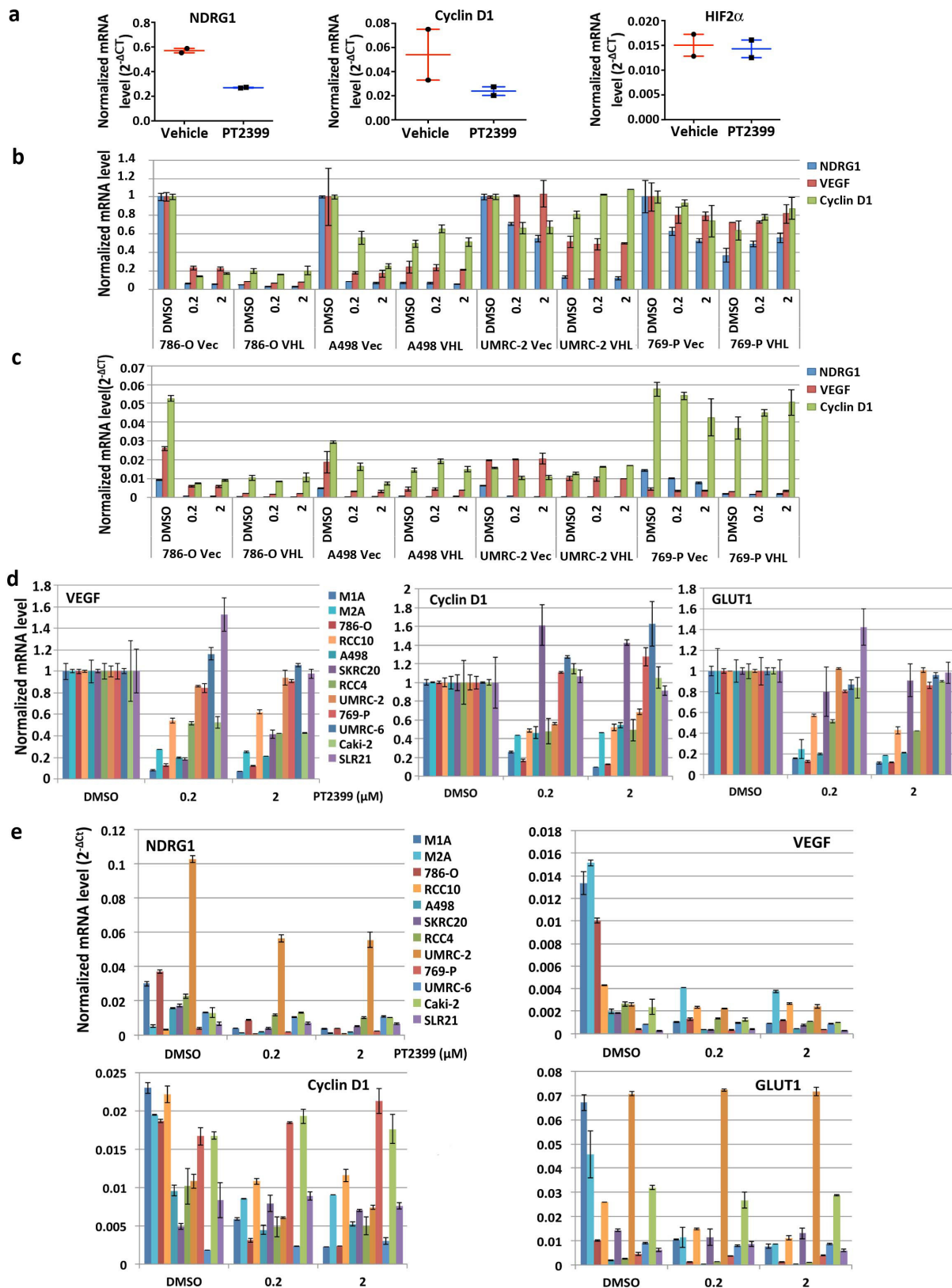
Extended Data Figure 6 | Antitumour activity of PT2399 using A498 cells. **a, b**, Representative BLI (**a**) and quantification of BLI measurements (**b**) of orthotopic tumours formed by A498 cells expressing firefly luciferase under the control of a CMV promoter before and after (30 days) treatment with PT2399 (30 mg kg $^{-1}$) or vehicle twice daily by oral gavage (vehicle, $n=10$; PT2399, $n=10$ mice from two independent experiments). **c, d**, Representative tumours (**c**) and tumour masses (**d**) at necropsy from mice treated as in **a** (vehicle, $n=10$; PT2399, $n=10$ mice). **e**, Serum

VEGF concentrations from mice treated as in **a** at time of necropsy (vehicle, $n=4$; PT2399, $n=4$ mice from two independent experiments). **f**, Immunoblot of representative tumours from **a**; vehicle, $n=4$; PT2399, $n=3$. **g**, Levels of the indicated mRNAs, normalized to *ACTB*, in A498 orthotopic tumours (vehicle, $n=4$; PT2399, $n=3$ mice from two independent experiments) treated as in **a**. Data shown as median with range (**b, d, e, g**). Statistical significance was assessed by using two-tailed Student's *t*-tests with Welch's correction (**b, d**) or Mann-Whitney test (**e**).



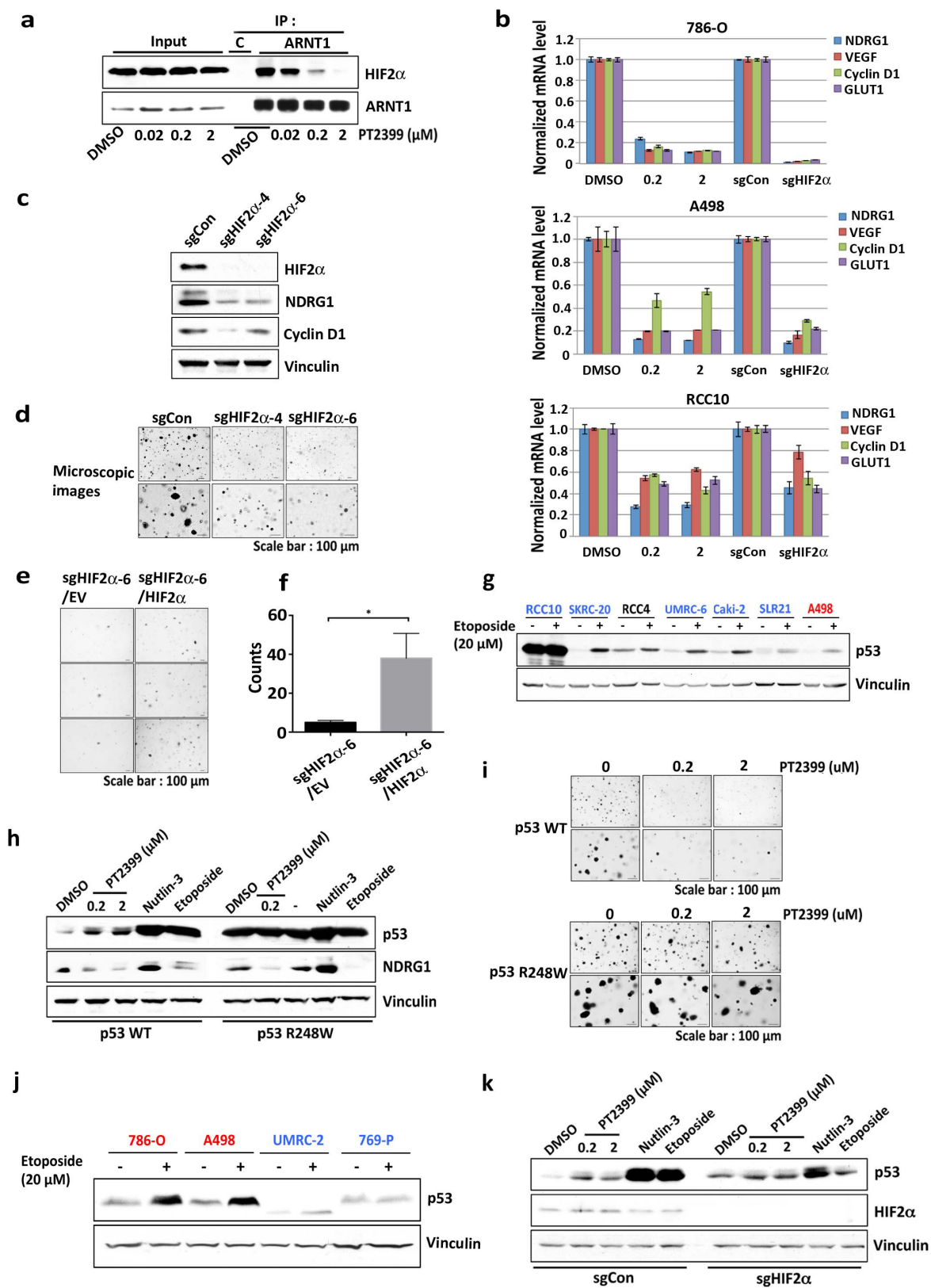
Extended Data Figure 7 | Elimination of HIF-1 α does not render UMRC-2 cells sensitive to PT2399 in soft agar assays. **a–c**, Firefly luciferase activity in the indicated cell lines after infection with a virus containing firefly luciferase under the control of a HIF-responsive (HRE-Luc) promoter (**a**, **c**) or CMV promoter (**b**) and treatment with the indicated concentrations of PT2399 for 16 h relative to DMSO-treated controls; $n = 3$ biological replicates. **d**, Immunoblot analysis of HRE-Luc-expressing UMRC-2 cells after CRISPR-based gene editing with control

sgRNA or HIF-1 α sgRNA (guides 2 and 3). Note that deletion of HIF-1 α in **c** and **d** was used to eliminate the contribution of HIF-1 α in **a**, **e**, **f**, Immunoblot (**e**) and mRNA levels (**f**) of UMRC-2 cells after CRISPR-based gene editing with control sgRNA or HIF-1 α sgRNA (guides 2 and 3). In **f** mRNA levels were normalized to *ACTB* and then to the corresponding control sgRNA value; $n = 3$ biological replicates. **g**, Soft agar assays of the cells analysed in **e** and **f** in the presence of the indicated concentrations of PT2399; $n = 3$ biological replicates. Data shown as mean \pm s.e.m. (**a–c**, **f**).



Extended Data Figure 8 | Variable sensitivity of ccRCC lines to PT2399 and pVHL. **a**, Levels of the indicated mRNAs, normalized to *ACTB*, in UMRC-2 orthotopic tumours treated with PT2399 (30 mg kg⁻¹) or vehicle twice daily for 1 month; vehicle, $n = 2$; PT2399, $n = 2$ mice from one experiment. **b**, **c**, Downregulation of HIF-responsive mRNAs by PT2399 in indicated cell lines. For each cell line the mRNA levels were normalized to *ACTB* mRNA (**c**) and then normalized to the untreated value for that cell line (**b**); $n = 3$ biological replicates. Data shown as median with

range (**a**) and mean \pm s.d. (**b**, **c**). **d**, **e**, Variable suppression of HIF target genes by PT2399 across a panel of ccRCC cell lines. Downregulation of HIF-responsive *VEGFA*, *CCND1* and *SLC2A1* mRNAs by PT2399 in the indicated cell lines. For each cell line the mRNA levels were normalized to *ACTB* (**e**) and then normalized to the untreated value for that cell line (**d**); $n = 3$ biological replicates. *VHL*^{+/+} SLR21 renal carcinoma cells were included for comparison. Data shown as mean \pm s.d.

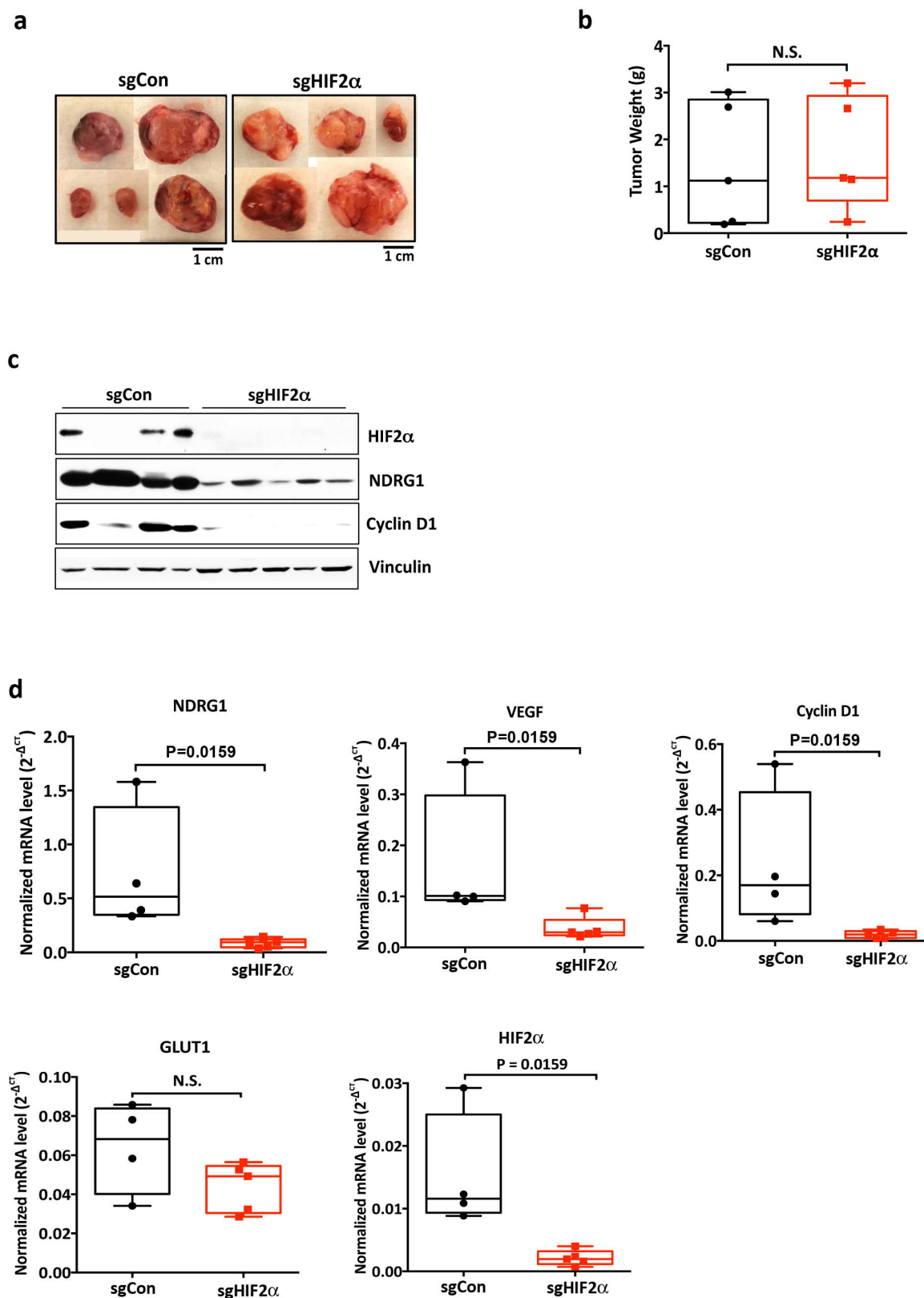


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | HIF-2 dependence of RCC10 cells.

a, Immunoblot analysis of anti-ARNT1 immunoprecipitates (IP) and whole cell extracts (input) prepared from RCC10 cells treated with increasing amounts of PT2399 or DMSO. **c**, control IP without ARNT1 antibody. **b**, Levels of the indicated mRNAs, normalized to *ACTB*, in 786-O, A498 and RCC10 cells treated with PT2399 at the indicated concentrations for 24 h or an effective HIF-2 α sgRNA (sgHIF-2 α -6), and then normalized to cells treated with DMSO or a control sgRNA, respectively. Data shown as mean \pm s.d.; $n = 3$ biological replicates. **c**, Immunoblot analysis of RCC10 cells after CRISPR-based editing with HIF-2 α sgRNAs or control sgRNA. **d**, **e**, soft agar colonies formed by RCC10 cells as in **c**; $n = 3$ biological replicates. In **e** cells were engineered to express an exogenous sgRNA-resistant HIF-2 α or empty vector (EV); $n = 3$. **f**, Soft agar colony counts as in **e** using ImageJ software. Colonies were counted using the following criteria: circularity range from 0.5 to 1.0 and size (pixels²) from 200 to infinity. Data shown as mean \pm s.e.m. Statistical significance was assessed by using two-tailed Student's *t*-tests

(**f**). * $P < 0.05$. **g–k**, p53 pathway status in ccRCC lines. **g**, **j**, Immunoblot analysis of the indicated cell lines treated for 16 h with etoposide or vehicle. Note overproduction of p53 in RCC10 cells and off-size p53 band in UMRC-2 cells. SLR21 cells are *VHL*^{+/+}. Red, PT2399 sensitive in soft agar assays. Blue, PT2399 insensitive. RCC4 cells do not form soft agar colonies and are therefore indeterminate. **h**, Immunoblot analysis of 786-O cells that were infected with an empty lentivirus conferring puromycin resistance and then later found to have spontaneously acquired a p53 mutation (R248W) compared to cells that retained wild-type p53. Cells were treated with PT2399 for 48 h or with nutlin-3 (30 μ M) or etoposide (20 μ M) for 16 h. **i**, Soft agar colony formation from cells in **h** treated with PT2399; $n = 3$ biological replicates. **k**, Immunoblot analysis of parental 786-O cells that underwent CRISPR-based gene editing with a control sgRNA or HIF-2 α sgRNA (guide 6) (as in Extended Data Fig. 2a) and were then treated with PT2399 for 58 h or treated with nutlin-3 (30 μ M) or etoposide (20 μ M) for 10 h.



Extended Data Figure 10 | Loss of HIF-2 α does not suppress UMRC-2 orthotopic tumour growth. **a–c**, Tumours (**a**), tumour weights (**b**), and tumour immunoblots (**c**) at necropsy from mice after orthotopic injection of UMRC-2 cells that had undergone CRISPR-based editing with control sgRNA or sgHIF-2 α -6 as in Fig. 5g; sgCon, $n = 5$; sgHIF-2 α , $n = 5$ mice from two independent experiments. The reason for the variable HIF-2 α

levels in **c** is unknown but could reflect, at least partly, variable numbers of host-derived cells in the tumour samples. **d**, Levels of the indicated mRNAs, normalized to *ACTB*, in tumours from **a–c**. Data shown as median with range (**b**, **d**). Statistical significance was assessed by using two-tailed Student's *t*-tests (**b**) or Mann–Whitney test (**d**). Loss of HIF-2 α did suppress subcutaneous tumour growth (data not shown).

Targeting renal cell carcinoma with a HIF-2 antagonist

Wenfang Chen^{1,2,3*}, Haley Hill^{1,2*}, Alana Christie^{1*}, Min Soo Kim^{1,4*}, Eboni Holloman^{1,2}, Andrea Pavia-Jimenez^{1,2}, Farrah Homayoun^{1,2}, Yuanqing Ma^{1,2}, Nirav Patel^{1,2}, Paul Yell⁵, Guiyang Hao⁶, Qurratulain Yousuf^{1,2}, Allison Joyce^{1,2}, Ivan Pedrosa^{1,6}, Heather Geiger⁷, He Zhang^{1,4}, Jenny Chang¹, Kevin H. Gardner^{8,9,10}, Richard K. Bruick^{1,11}, Catherine Reeves⁷, Tae Hyun Hwang^{1,4}, Kevin Courtney^{1,2}, Eugene Frenkel^{1,2}, Xiankai Sun^{1,6}, Naseem Zojwalla¹², Tai Wong¹², James P. Rizzi¹², Eli M. Wallace¹², John A. Josey¹², Yang Xie^{1,4}, Xian-Jin Xie^{1,4}, Payal Kapur^{1,13}, Renée M. McKay^{1,2} & James Brugarolas^{1,2}

Clear cell renal cell carcinoma (ccRCC) is characterized by inactivation of the von Hippel-Lindau tumour suppressor gene (*VHL*)^{1,2}. Because no other gene is mutated as frequently in ccRCC and *VHL* mutations are truncal³, *VHL* inactivation is regarded as the governing event⁴. *VHL* loss activates the HIF-2 transcription factor, and constitutive HIF-2 activity restores tumorigenesis in *VHL*-reconstituted ccRCC cells⁵. HIF-2 has been implicated in angiogenesis and multiple other processes^{6–9}, but angiogenesis is the main target of drugs such as the tyrosine kinase inhibitor sunitinib¹⁰. HIF-2 has been regarded as undruggable¹¹. Here we use a tumourgraft/patient-derived xenograft platform^{12,13} to evaluate PT2399, a selective HIF-2 antagonist that was identified using a structure-based design approach. PT2399 dissociated HIF-2 (an obligatory heterodimer of HIF-2 α -HIF-1 β)¹⁴ in human ccRCC cells and suppressed tumorigenesis in 56% (10 out of 18) of such lines. PT2399 had greater activity than sunitinib, was active in sunitinib-progressing tumours, and was better tolerated. Unexpectedly, some *VHL*-mutant ccRCCs were resistant to PT2399. Resistance occurred despite HIF-2 dissociation in tumours and evidence of Hif-2 inhibition in the mouse, as determined by suppression of circulating erythropoietin, a HIF-2 target¹⁵ and possible pharmacodynamic marker. We identified a HIF-2-dependent gene signature in sensitive tumours. Gene expression was largely unaffected by PT2399 in resistant tumours, illustrating the specificity of the drug. Sensitive tumours exhibited a distinguishing gene expression signature and generally higher levels of HIF-2 α . Prolonged PT2399 treatment led to resistance. We identified binding site and second site suppressor mutations in HIF-2 α and HIF-1 β , respectively. Both mutations preserved HIF-2 dimers despite treatment with PT2399. Finally, an extensively pretreated patient whose tumour had given rise to a sensitive tumourgraft showed disease control for more than 11 months when treated with a close analogue of PT2399, PT2385. We validate HIF-2 as a target in ccRCC, show that some ccRCCs are HIF-2 independent, and set the stage for biomarker-driven clinical trials.

The discovery of a 280 Å³ cavity within the PAS-B domain of HIF-2 α ^{16,17} and subsequent identification of compounds that bound this cavity and dissociated HIF-2 α from HIF-1 β ¹⁸ led to an iterative structure-based program that identified selective, potent HIF-2 α antagonists such as PT2399 (described in ref. 19) and PT2385 (ref. 20).

To evaluate PT2399 in renal cancer, we tested a panel of 22 independently generated tumourgrafts^{12,13} (Extended Data Table 1). To assess the drug's tolerability, we evaluated its effects on weight and blood counts in mice bearing these tumourgrafts. PT2399 did not induce weight loss, whereas sunitinib, at doses matching human exposures¹³, did (Fig. 1a). However, PT2399 caused modest anaemia and leukopenia (Fig. 1b and Extended Data Fig. 1a).

We hypothesized that the reduction in haemoglobin (2.0 g dl⁻¹; $P=0.0001$) was due to a decrease in erythropoietin (EPO), which is regulated by HIF-2¹⁵. Consistent with this hypothesis, the number of red blood cell precursors was decreased by 35% ($P<0.0001$; Fig. 1b) and the level of EPO, which may serve as a pharmacodynamic marker, was suppressed by 75% ($P<0.0001$; Fig. 1b).

PT2399 decreased tumour growth by 60% across all tumourgrafts evaluated ($P<0.0001$; Fig. 1c). According to their responsiveness, tumourgrafts were classified into sensitive (tumour growth inhibition at last measurement $>80\%$), intermediate (40–80%), and resistant ($<40\%$; Extended Data Table 1). Forty-five percent of tumourgrafts were sensitive (10/22), 23% intermediate, and 32% resistant (Fig. 1d and Extended Data Fig. 1b, c). Sensitive tumours included tumours with aggressive sarcomatoid and rhabdoid features (Extended Data Table 1). Among ccRCC tumourgrafts, 56% (10/18) were sensitive. Unexpectedly, four ccRCCs were resistant to PT2399, including three with *VHL* mutations (Extended Data Table 1).

PT2399 was more active than sunitinib ($P=0.0126$) (Fig. 1c and Extended Data Fig. 1b) and inhibited tumour growth in several sunitinib-resistant tumours (Fig. 1d). There was no bias in treatment allocation, as treatment groups were balanced (pre-trial: tumour size, $P=0.11$; tumour growth rate, $P=0.22$; and mouse weight, $P=0.34$). PT2399 reduced tumour cell density and increased fibrosis (Extended Data Fig. 1c–e). Ki67 immunohistochemistry (IHC) showed that PT2399 inhibited tumour cell proliferation by 3.5-fold (mean value change of -19.5 ± 2.4 ; $P<0.0001$; Extended Data Fig. 1e, f). Inhibition of cell proliferation was also observed in live mice using 3'-[¹⁸F] fluoro-3'-deoxythymidine positron emission tomography/computerized tomography (PET/CT) scanning (Extended Data Fig. 1g, h). In addition, PT2399 collapsed the tumour vasculature, decreasing vascular area threefold (mean value change of -29.1 ± 6.1 ; $P=0.0011$) (Extended Data Fig. 1e, f). To determine whether changes in vascular area were due to inhibition of tumour VEGF, we exploited the species difference between graft (human) and host (mouse). PT2399

¹Kidney Cancer Program, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ²Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ³Department of Pathology, the First Affiliated Hospital of Sun Yat-sen University, Guangzhou, 510080, People's Republic of China.

⁴Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁵Parkland Health and Hospital System, Dallas, Texas 75235, USA. ⁶Department of Radiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁷New York Genome Center, New York, New York 10013, USA. ⁸Structural Biology Initiative, CUNY Advanced Science Research Center, New York, New York 10031, USA. ⁹Department of Chemistry and Biochemistry, City College of New York, New York, New York 10031, USA. ¹⁰Biochemistry, Chemistry and Biology Ph.D. Programs, Graduate Center, City University of New York, New York, New York 10016, USA. ¹¹Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ¹²Peloton Therapeutics Inc., Dallas, Texas 75235, USA. ¹³Department of Pathology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

*These authors contributed equally to this work.

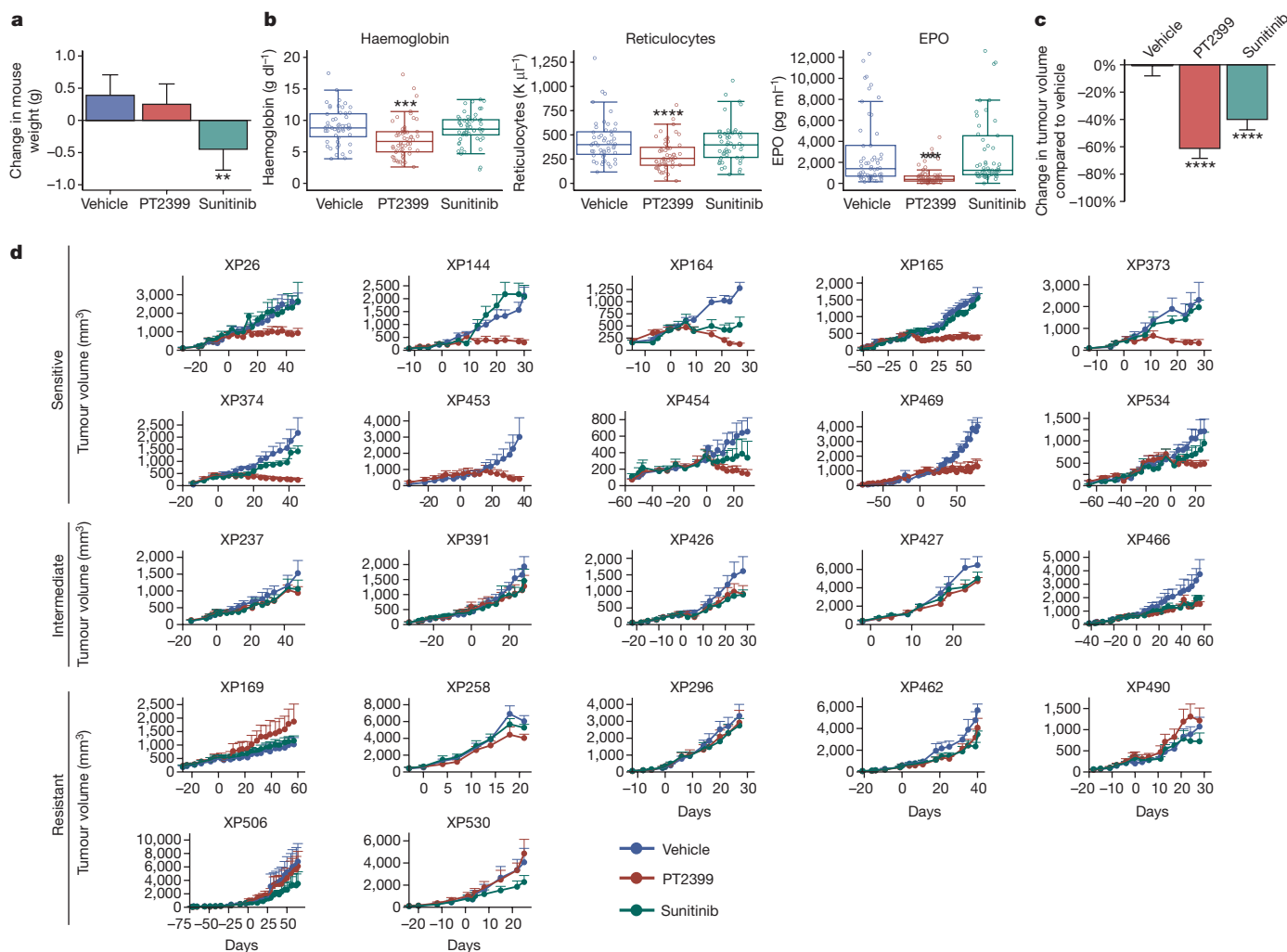


Figure 1 | Evaluation of PT2399 in RCC tumourgraft-bearing mice.

a, Mean change in mouse body weights after treatment with vehicle ($n = 89$), PT2399 (100 mg kg^{-1}) by oral gavage every 12 h ($n = 96$) or sunitinib (10 mg kg^{-1}) by oral gavage every 12 h ($n = 82$). **b**, Haemoglobin levels, reticulocyte counts, and erythropoietin (EPO) levels in mice treated as indicated. Haemoglobin and reticulocytes: vehicle $n = 52$ mice, PT2399 $n = 58$, sunitinib $n = 53$; EPO: vehicle $n = 63$, PT2399 $n = 74$, sunitinib $n = 61$. **c**, Mean per cent change in tumour volume in mice treated with vehicle ($n = 89$), PT2399 ($n = 96$), or sunitinib ($n = 82$). **d**, Growth curves of each tumourgraft line grouped according to PT2399 responsiveness into sensitive (growth inhibition (GI) at end of trial $> 80\%$), intermediate (GI = $40\text{--}80\%$), or resistant (GI $< 40\%$). Treatment starts on day 0

suppressed circulating human VEGF by 93%, but mouse VEGF was unaffected (Extended Data Fig. 1i). Thus, tumour VEGF production, but not extratumoural VEGF, is HIF-2-dependent and inhibited by PT2399. This tumour selectivity represents a marked improvement over current angiogenesis inhibitors. PT2399 also inhibited VEGF production in tumours progressing on sunitinib (Extended Data Fig. 2).

We evaluated the effects of PT2399 on HIF-2 in tumours. Immunoprecipitation of the HIF-1 β subunit, which is shared by both HIF-2 α and HIF-1 α , showed that PT2399 specifically disassembled HIF-2 but not HIF-1 complexes (Fig. 2a). Similar results were observed using a proximity ligation assay (Fig. 2b). Correspondingly, PT2399 reduced the expression of HIF-2 target genes (*VEGFA*, *SERPINE1* (encoding PAI-1), *IGFBP3*, *CCND1* (encoding cyclin D1), *TGFA*, and *SLC2A1* (encoding GLUT1); all comparisons $P < 0.05$; Fig. 2c), but not HIF-1 targets (*CA9*, *PGK1*, and *LDHA*).

Notably, PT2399 did not affect the majority of HIF-2 target genes in resistant tumours (Fig. 2c). A modest decrease in *VEGFA* mRNA

and values and error bars represent mean tumour volume \pm s.e.m. To minimize bias (despite overestimation) volumes were calculated as length \times width \times height. Each tumourgraft line had $n \approx 3\text{--}5$ tumours per treatment group (vehicle $n = 89$ mice, PT2399 $n = 96$, sunitinib $n = 82$). **a–c**, Tests completed using a mixed model with compound symmetrical covariance structure for mice in the same tumourgraft line using vehicle as the reference group. Reticulocyte values were log-transformed for analysis; EPO levels were Box-Cox transformed; raw values are depicted in all figures. All bar charts depict the mean tumour volume with the error bar representing s.e.m., while all boxplots have median centre values. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

did not translate into lower circulating vascular endothelial growth factor (VEGF; Fig. 2d). However, as determined by a reduction in EPO ($P = 0.0002$; Fig. 2d), Hif-2 was inhibited by PT2399 in mice with resistant tumours. Furthermore, immunoprecipitation experiments showed that HIF-2 complexes were dissociated in resistant tumours (Fig. 2a). Thus, somewhat unexpectedly, PT2399 disassembled HIF-2 in resistant tumours, but HIF-2 target genes were largely unaffected.

To better characterize the effects of PT2399, we performed RNA sequencing (RNA-seq) on 46 tumours (Extended Data Tables 1, 2). In sensitive tumours, we identified 492 RNAs that were deregulated by PT2399 (FDR < 0.05 ; Fig. 2e and Extended Data Table 3). By contrast, the same analysis in resistant tumours found no genes that were deregulated by PT2399 (Fig. 2e, f). Similar results were obtained by an independent, blinded analysis (H.G. and C.R.). The selective changes induced by PT2399 in sensitive tumours suggest that PT2399 sensitivity is linked to its ability to alter gene expression. Furthermore, the lack of gene expression changes in resistant tumours suggest that PT2399

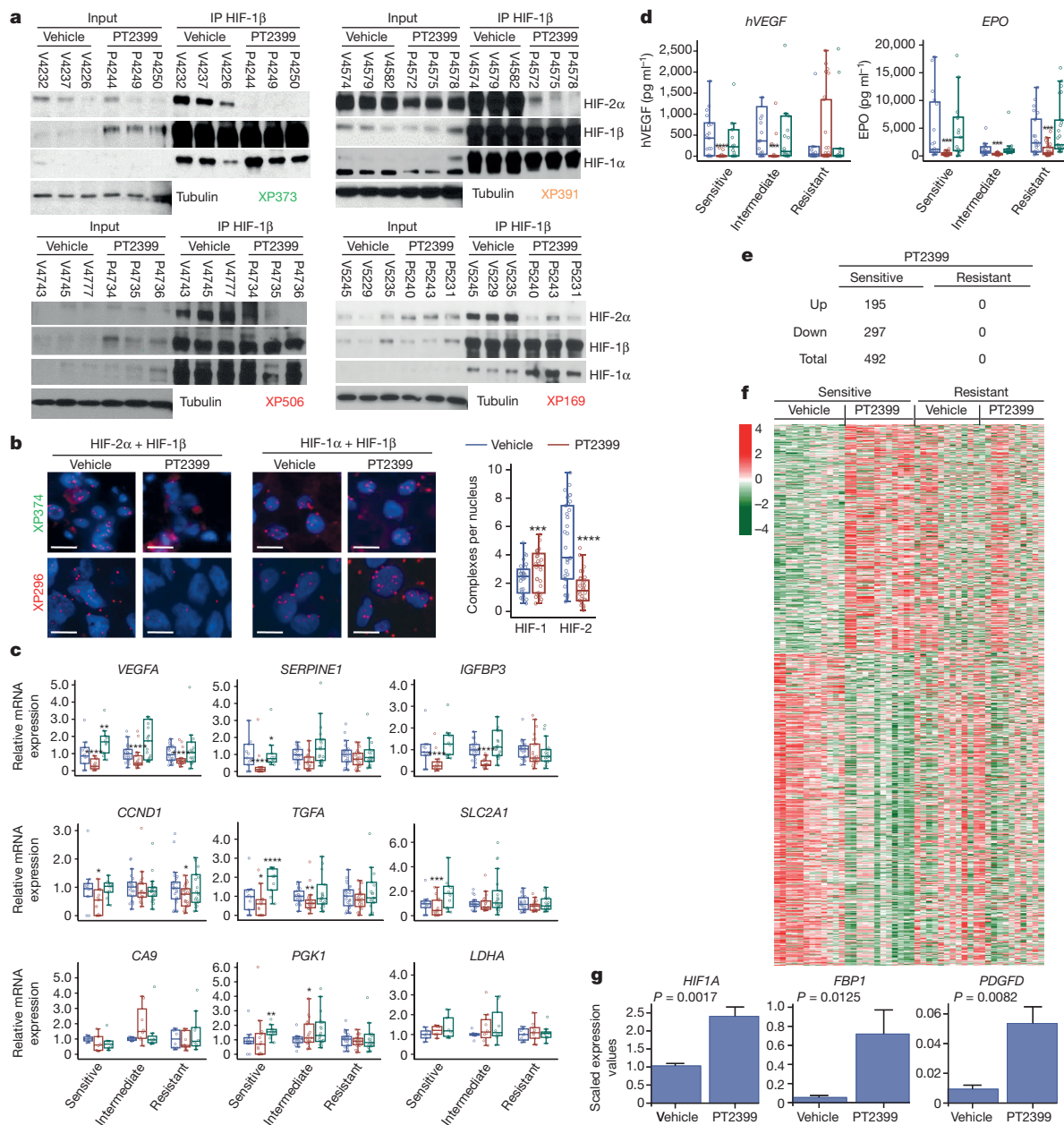


Figure 2 | PT2399 dissociates HIF-2 complexes in sensitive and resistant RCCs and induces changes in gene expression in sensitive tumours.

a, Immunoprecipitation of HIF-1 β from tumour lysates of sensitive (XP373), intermediate (XP391), and resistant (XP506 and XP169) tumours from mice treated with vehicle or PT2399. Samples are labelled with V for vehicle-treated or P for PT2399-treated followed by the mouse identifier.

b, Proximity ligation assay detecting HIF-2 α -HIF-1 β or HIF-1 α -HIF-1 β heterodimers from vehicle- or PT2399-treated sensitive (XP374) or resistant (XP296) tumours and summary of results across responsive and resistant tumourgrafts. Images representative of quantitative data shown in graph. Summary includes analyses from 11 vehicle-treated tumours and 11 PT2399-treated tumours (3 fields were analysed for each sample) in 5 sensitive, 3 intermediate, and 3 resistant tumourgraft trials. Scale bars, 20 μ m.

c, qRT-PCR for the indicated HIF-2 target genes in PT2399-sensitive, -intermediate, and -resistant tumours treated with vehicle (blue), PT2399 (red), or sunitinib (green). HIF-1 target genes *CA9*, *PGK1*, and *LDHA* included as negative controls. RT-PCR was repeated three times for each sample. Except for *PGK1* and *LDHA*, samples were available for $n = 58$ vehicle-treated tumours (sensitive: $n = 11$; intermediate: $n = 21$; resistant: $n = 26$), $n = 62$ PT2399-treated tumours (sensitive: $n = 15$; intermediate: $n = 21$; resistant: $n = 26$), and $n = 52$ sunitinib-treated tumours (sensitive: $n = 10$; intermediate: $n = 23$; resistant: $n = 19$). *PGK1* and *LDHA* data were available for 24 tumours for each treatment group (sensitive: $n = 6$;

intermediate: $n = 8$; resistant: $n = 10$). **d**, Circulating tumour-produced hVEGF and mouse EPO levels in mice with sensitive, intermediate, and resistant tumours treated with vehicle (blue), PT2399 (red), and sunitinib (green). Enzyme-linked immunosorbent assay (ELISA) data were generated for 63 vehicle-treated tumours (sensitive: $n = 21$; intermediate: $n = 19$; resistant: $n = 23$), 74 PT2399-treated tumours (sensitive: $n = 27$; intermediate: $n = 21$; resistant: $n = 26$), and 61 sunitinib-treated tumours (sensitive: $n = 15$; intermediate: $n = 23$; resistant: $n = 23$). **e**, Number of RNAs upregulated and downregulated by PT2399 in sensitive and resistant tumours. **f**, Heatmap representation from RNA-seq analysis showing genes differentially regulated by PT2399 in sensitive and resistant tumours. Removal of an unclassified tumour (XP169) from the resistant group did not affect conclusions. **g**, RNA-seq analyses showing increased expression of selected genes by PT2399 in sensitive tumours. **b–d, g**, Tests completed using a mixed model with compound symmetrical covariance structure for mice in the same tumourgraft line using vehicle as the reference group. qRT–PCR levels were log-transformed for analysis; EPO and hVEGF levels were Box-Cox transformed; RNA-seq levels were \log_2 -transformed; raw values depicted in all graphs. All bar charts depict the mean with the error bar representing s.e.m., while all boxplots have median centre values. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$. See Supplementary Fig. 1 for gel source images.

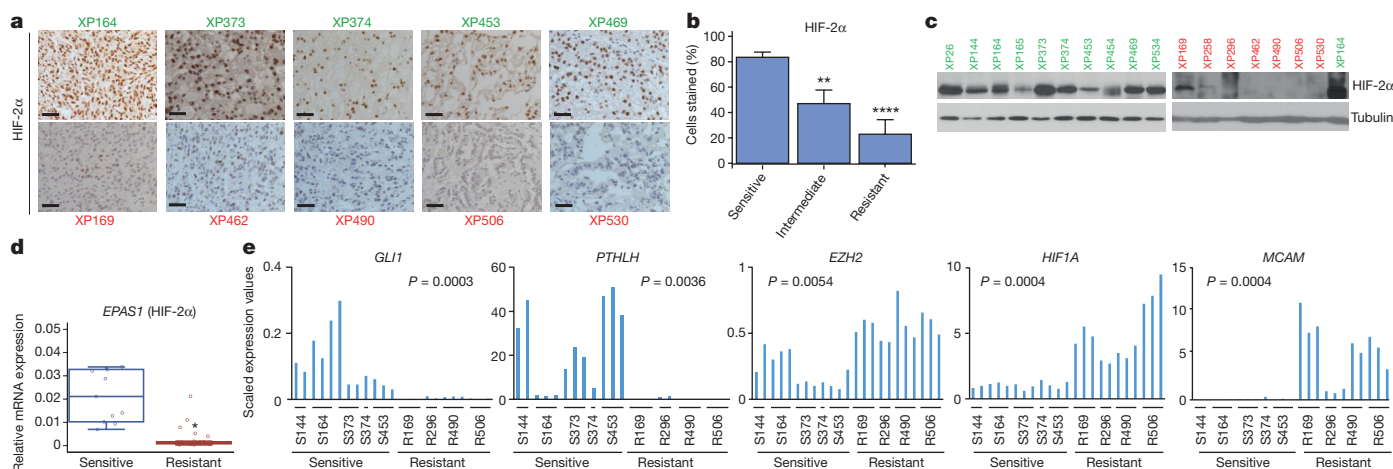


Figure 3 | Sensitive and resistant tumours can be distinguished by HIF-2α levels and gene expression signature. **a**, HIF-2α expression by immunohistochemistry (IHC) in sensitive (green) and resistant (red) tumours. Scale bars, 50 μm. Images representative of quantitative data shown in **b**. **b**, Quantification of HIF-2α-positive cells as determined by IHC in sensitive, intermediate, and resistant tumours from all 22 tumourgraft lines (sensitive: $n = 10$; intermediate: $n = 5$; resistant: $n = 7$). **c**, Western blot analysis of sensitive (green) and resistant (red) tumourgraft lines. XP164 lysate loaded twice as a reference for comparison between the two membranes. **d**, qRT-PCR of EPAS1 (HIF-2α) expression in sensitive ($n = 11$) versus resistant ($n = 26$) vehicle-treated tumourgrafts. RT-PCR

was repeated three times for each sample. **e**, Candidate genes from RNA-seq analysis differentially expressed in sensitive and resistant tumours. **f**, ANOVA used to determine whether sensitive tumours were different from intermediate or resistant. Bar chart depicts the mean with the error bar representing s.e.m. **d**, **e**, Tests completed using a mixed model analysis with compound symmetrical covariance structure for mice in the same tumourgraft line. RNA-seq values were log₂-transformed for analysis; raw values depicted in all graphs. Bar charts depict individual RNA-seq values, while all boxplots have median centre values. * $P < 0.05$; ** $P < 0.01$; **** $P < 0.0001$. See Supplementary Fig. 1 for gel source images.

is highly specific. Consistent with PT2399 specificity, PT2399 had less effect on overall gene expression than did subtle differences among patients' tumours (Extended Data Fig. 3a).

Extensive studies have investigated HIF-2 target genes in ccRCC^{21–23}. However, by leveraging (i) PT2399 specificity; (ii) RCC tumourgrafts, with minimal human stroma²⁴; and (iii) an RNA-seq algorithm excluding contaminating mouse (stromal) transcripts, we were able to define the HIF-2 program particularly accurately. Among the 492 RNAs that were deregulated in PT2399-sensitive tumours, 439 were protein coding, and 271 were downregulated; these included previously identified canonical HIF-2 targets (*IGFBP3*, *SERPINE1*, *VEGFA*, *CCND1*) as well as other genes such as *LOX*, *CXCR4*, *IL6* and *DDIT4* (also known as *REDD1*) (Extended Data Fig. 3b). Pathway and gene set enrichment analyses showed downregulation of cell cycle, DNA replication, cell cycle checkpoint, and DNA repair processes (Extended Data Table 4). Regulation of DNA repair genes by HIF-2, as previously observed in cell lines⁶, may explain the resistance of ccRCC to radiotherapy. PT2399 increased the expression of 168 protein-coding genes, including fibrosis-related genes (i.e. *PDGFD*), *HIF1A* (previously shown to be induced by HIF-2α knockdown²²), and *FBP1*, a gluconeogenic gene recently reported to suppress RCC progression²⁵ (Fig. 2g and Extended Data Table 4).

We sought to identify a biomarker that could distinguish between PT2399-sensitive and resistant tumours. We found that HIF-2α protein was expressed in 83% of cells in sensitive tumours compared to 23% in resistant tumours ($P < 0.0001$; Fig. 3a, b and Extended Data Fig. 4a). Although there were differences even within tumours, higher HIF-2α expression in sensitive tumours was observed by western blotting (Fig. 3c and Extended Data Fig. 4b) and RT-PCR (Fig. 3d). Lower, and at times undetectable, HIF-2α levels in resistant tumours may explain why PT2399 does not affect gene expression in this group.

Next, we compared RNA-seq data sets between sensitive and resistant vehicle-treated tumours. Using a rigorous Wilcoxon test, we identified 1,327 differentially expressed RNAs (Extended Data Table 3), including 94 (76 mRNAs) that were uniformly over- or underexpressed across every sensitive versus resistant tumour sample (Extended Data Fig. 4c and Extended Data Table 3). *GLI1*, a transcription factor of the sonic

hedgehog family, and *PTH1H* (parathyroid hormone-like hormone), a neuroendocrine peptide that has been implicated in epithelial-mesenchymal interactions and calcium ion transport, were uniformly overexpressed in sensitive tumours (Fig. 3e). Notably, *HIF1A* expression was increased in the resistant group (Fig. 3e). Increased expression of HIF-1α protein was also observed by immunohistochemistry in some, but not all, resistant tumours (Extended Data Fig. 4a). *EZH2* and *MCAM* were also uniformly overexpressed in resistant tumours (Fig. 3e).

Overall, our data show that ccRCC can be classified into HIF-2-dependent and -independent tumours, and that these tumours differ in HIF-2α (and possibly HIF-1α) levels and in their baseline gene expression. These tumour subtypes did not correlate with BRCA1 associated protein 1 (BAP1) and polybromo 1 (PBRM1) status²⁶ in this small series (Extended Data Table 1). Our results point to different mechanisms of tumorigenesis downstream of VHL that may underlie differences in tumour behaviour or responsiveness to therapy.

Given the differences in gene expression, we investigated whether sensitive and resistant tumours showed differing imaging characteristics. We obtained CT scan images from patient tumours giving rise to tumourgrafts before surgery. The sensitive group was characterized by tumours with peripheral hypervascularity and a central non-enhancing area (typical of high-grade ccRCC²⁷) and, if present, tumour infiltration was focal (Extended Data Fig. 5). The resistant group was more heterogeneous, but several tumours were relatively hypovascular and diffusely infiltrating (Extended Data Fig. 5).

We investigated whether sensitive tumours would acquire resistance. We exposed mice bearing tumours formed from a sensitive tumourgraft (XP164) to prolonged treatment with PT2399 or sunitinib. Sunitinib resistance developed within 60 days (Fig. 4a; compare to Fig. 1d), but resistance to PT2399 took > 120 days (Fig. 4a). PT2399 resistance was associated with increased tumour vascularity and higher tumour VEGF production (Fig. 4b). We sequenced the HIF-2α gene (*EPAS1*) and identified a c.968G > A heterozygous mutation resulting in a G323E substitution in one tumour (Fig. 4c). The mutation was absent in a vehicle-treated tumour and in the second resistant tumour

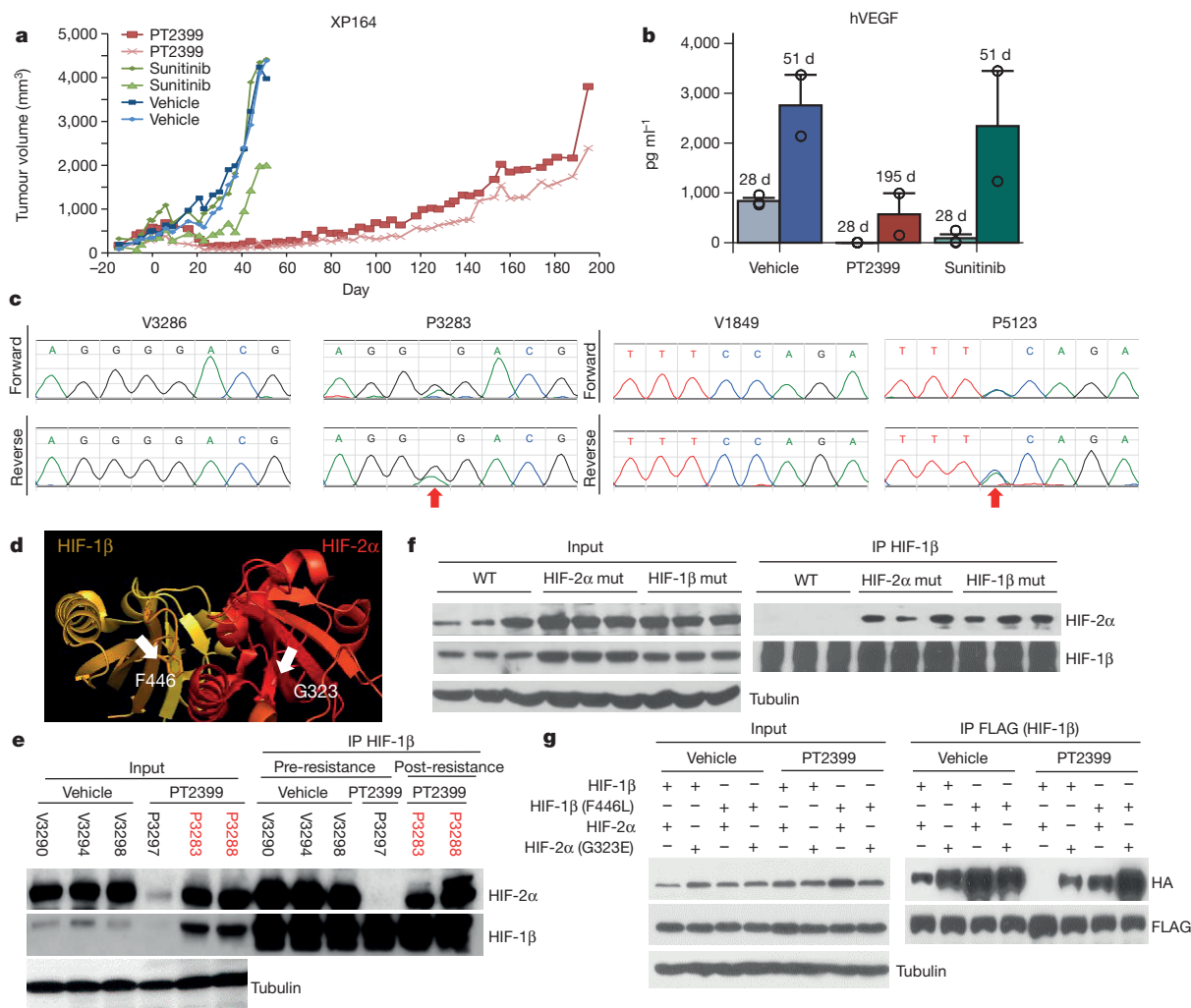


Figure 4 | Acquired resistance following prolonged PT2399 exposure.

a, Tumour volumes from a cohort of mice of the XP164 tumourgraft line treated with vehicle (blue lines, $n = 2$; V3286 and V3299); sunitinib until the development of resistance (green lines, $n = 2$; S3295 and S3296; compare to Fig. 1d); or PT2399 (red lines, $n = 2$; P3283 and P3288). **b**, Circulating human VEGF levels in mice treated for the indicated number of days (d) showing increased tumour-produced VEGF with development of resistance (28 day bars, $n = 3$; all other bars, $n = 2$). The bar chart depicts the mean with the error bar representing s.e.m. **c**, Bidirectional chromatograms from tumourgrafts that developed resistance compared to controls: P3283 (c.968G > A in *EPAS1* (*HIF2A*) leading to G323E) and P5123 (derived from P3288; c.1338C > A in *ARNT* (*HIF1B*) leading to F446L). **d**, Crystal structure of PAS-B domains from HIF-2 α bound to HIF-1 β (PDB entry 4ZP4; ref. 28) highlighting side

chains of G323 (lining opening of PT2399 binding pocket in HIF-2 α) and F446 (in HIF-1 β at the interface with HIF-2 α). In another structure (PDB entry 4GHI; ref. 18) the quaternary arrangement between HIF-2 α and HIF-1 β PAS-B domains differs, but F446 remains at the interface. **e**, HIF-1 β immunoprecipitation from XP164 tumourgrafts before and after (red) development of resistance showing reformation of HIF-2 α –HIF-1 β complexes following the acquisition of resistance (V, vehicle; P, PT2399). **f**, HIF-1 β immunoprecipitation from tumours of mice with HIF-2 α or HIF-1 β mutations (or wild-type controls) treated with PT2399 ($n = 3$ mice per group). **g**, FLAG immunoprecipitation from HEK293T cells transfected with plasmids encoding FLAG-tagged HIF-1 β (FLAG–HIF-1 β ; FLAG–HIF-1 β –F446L) or HA-tagged HIF-2 α (HA–HIF-2 α ; HA–HIF-2 α –G323E) treated with vehicle or PT2399. See Supplementary Fig. 1 for gel source images.

(despite originating from the same parental tumour). Structural analyses of HIF-2 α ^{17,28} showed that G323 is at the entrance to the cavity, where PT2399 binds (Fig. 4d). Akin to engineered mutations^{19,29}, a glutamate side chain would prevent PT2399 access. Consistent with this notion, PT2399 failed to dissociate HIF-2 complexes in mutant tumours (Fig. 4e).

We then sequenced HIF-1 β from the second resistant tumour, and failed to identify a mutation. Nevertheless, HIF-2 complexes had reformed (Fig. 4e). The tumour was passaged in mice, which were maintained on PT2399, and remained resistant. Immunoprecipitation experiments again showed dimeric HIF-2 complexes (Fig. 4f). Sequencing of passaged tumours revealed a heterozygous c.1338C > A mutation resulting in a F446L substitution in the HIF-1 β PAS-B domain (Fig. 4c). F446 is at the interface between HIF-1 β and HIF-2 α (Fig. 4d). We postulate that F446L functions as a second-site suppressor mutation

and that a more flexible side chain at the complex interface accommodates conformational changes induced by PT2399, allowing drug-bound HIF-2 α to bind to HIF-1 β . Experimentally, both HIF-1 β (F446L) and HIF-2 α (G323E), when expressed in cells, were sufficient to preserve the formation of HIF-2 dimers despite treatment with PT2399, and their effects appeared additive (Fig. 4g). Overall, these results pave the way for second generation inhibitors or complementary approaches to leverage other potential drug-binding pockets that have been recently revealed²⁸.

Finally, a patient with metastatic ccRCC, whose tumour gave rise to a sensitive tumourgraft (XP165), enrolled in a phase 1 trial with PT2385³⁰ (NCT02293980). The patient, a 47-year-old male, had originally presented with omental and abdominal wall metastases following a radical nephrectomy of a stage III, high-grade ccRCC. After a failed attempt at surgical removal, he had received high-dose IL2, bevacizumab,

sorafenib, everolimus, sunitinib, pazopanib, and axitinib. Despite extensive pretreatment, he remained free of progression on PT2385 for more than 11 months (Extended Data Fig. 6). These data validate HIF-2 as a target for ccRCC, provide insight into HIF-2-mediated tumorigenesis, establish variable tumour dependency on HIF-2 identifying different ccRCC subtypes and associated biomarkers that may be incorporated in future clinical trials, showcase the specificity of PT2399, and anticipate mechanisms of resistance.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 September 2015; accepted 26 August 2016.

Published online 5 September 2016.

- Gnarra, J. R. *et al.* Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nat. Genet.* **7**, 85–90 (1994).
- Herman, J. G. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl Acad. Sci. USA* **91**, 9700–9704 (1994).
- Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
- Kaelin, W. G., Jr & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell* **30**, 393–402 (2008).
- Kondo, K., Klcio, J., Nakamura, E., Lechpammer, M. & Kaelin, W. G., Jr Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer Cell* **1**, 237–246 (2002).
- Bertout, J. A. *et al.* HIF2 α inhibition promotes p53 pathway activity, tumor cell death, and radiation responses. *Proc. Natl Acad. Sci. USA* **106**, 14391–14396 (2009).
- Covello, K. L. *et al.* HIF-2 α regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes Dev.* **20**, 557–570 (2006).
- Ema, M. *et al.* A novel bHLH-PAS factor with close sequence similarity to hypoxia-inducible factor 1 α regulates the VEGF expression and is potentially involved in lung and vascular development. *Proc. Natl Acad. Sci. USA* **94**, 4273–4278 (1997).
- Gordan, J. D., Bertout, J. A., Hu, C. J., Diehl, J. A. & Simon, M. C. HIF-2 α promotes hypoxic cell proliferation by enhancing c-myc transcriptional activity. *Cancer Cell* **11**, 335–347 (2007).
- Mendel, D. B. *et al.* *In vivo* antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clin. Cancer Res.* **9**, 327–337 (2003).
- Koehler, A. N. A complex task? Direct modulation of transcription factors with small molecules. *Curr. Opin. Chem. Biol.* **14**, 331–340 (2010).
- Pavia-Jiménez, A., Tcheuyap, V. T. & Brugarolas, J. Establishing a human renal cell carcinoma tumorigraft platform for preclinical drug testing. *Nat. Protocols* **9**, 1848–1859 (2014).
- Sivanand, S. *et al.* A validated tumorigraft model reveals activity of dovitinib against renal cell carcinoma. *Sci. Transl. Med.* **4**, 137ra75 (2012).
- Tian, H., McKnight, S. L. & Russell, D. W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **11**, 72–82 (1997).
- Rankin, E. B. *et al.* Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin *in vivo*. *J. Clin. Invest.* **117**, 1068–1077 (2007).
- Erbel, P. J., Card, P. B., Karakuzu, O., Bruick, R. K. & Gardner, K. H. Structural basis for PAS domain heterodimerization in the basic helix-loop-helix-PAS transcription factor hypoxia-inducible factor. *Proc. Natl Acad. Sci. USA* **100**, 15504–15509 (2003).
- Scheuermann, T. H. *et al.* Artificial ligand binding within the HIF2 α PAS-B domain of the HIF2 transcription factor. *Proc. Natl Acad. Sci. USA* **106**, 450–455 (2009).
- Scheuermann, T. H. *et al.* Allosteric inhibition of hypoxia inducible factor-2 with small molecules. *Nat. Chem. Biol.* **9**, 271–276 (2013).
- Cho, H. *et al.* On-target efficacy of a HIF2 α antagonist in preclinical kidney cancer models. *Nature* <http://dx.doi.org/10.1038/nature19795> (2016).
- Wallace, E. M. *et al.* A small-molecule antagonist of HIF-2 α is efficacious in preclinical models of renal cell carcinoma. *Cancer Res.* **76**, 5491–5500 (2016).
- Hu, C. J. *et al.* Differential regulation of the transcriptional activities of hypoxia-inducible factor 1 α (HIF-1 α) and HIF-2 α in stem cells. *Mol. Cell. Biol.* **26**, 3514–3526 (2006).
- Raval, R. R. *et al.* Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. *Mol. Cell. Biol.* **25**, 5675–5686 (2005).
- Keith, B., Johnson, R. S. & Simon, M. C. HIF1 α and HIF2 α : sibling rivalry in hypoxic tumour growth and progression. *Nat. Rev. Cancer* **12**, 9–22 (2011).
- Peña-Llopis, S. *et al.* BAP1 loss defines a new class of renal cell carcinoma. *Nat. Genet.* **44**, 751–759 (2012).
- Li, B. *et al.* Fructose-1,6-bisphosphatase opposes renal carcinoma progression. *Nature* **513**, 251–255 (2014).
- Brugarolas, J. Molecular genetics of clear-cell renal cell carcinoma. *J. Clin. Oncol.* **32**, 1968–1976 (2014).
- Pedrosa, I. *et al.* MR classification of renal masses with pathologic correlation. *Eur. Radiol.* **18**, 365–375 (2008).
- Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303–308 (2015).
- Rogers, J. L. *et al.* Development of inhibitors of the PAS-B domain of the HIF-2 α transcription factor. *J. Med. Chem.* **56**, 1739–1747 (2013).
- Courtney, K. D. *et al.* A phase I dose escalation trial of PT2385, a first-in-class oral HIF-2 α inhibitor, in patients with advanced clear cell renal cell carcinoma. *J. Clin. Oncol.* **34**, Suppl. Abstract 2506 (2016).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the patients who generously provided tissues and participated in our studies. PT2399 was provided by Peloton Therapeutics, Inc. Funding was provided by Peloton Therapeutics, Inc. (OTD-105466), CPRIT (RP160440) and philanthropy, including the Tom Green memorial. W.C. is supported by grants from the National Natural Science Foundation of China (No. 811011934) and the Science and Technology Program of Guangzhou, China (No. 2012J5100031). M.S.K. and H.Z. are supported by a grant from CPRIT (RP150596). I.P. is supported by grants from the NIH (R01CA154475, P50CA196516). X.S. is supported by a grant from CPRIT (RP110771). J.B. is a Virginia Murchison Linthicum endowed scholar and is supported by grants from the NIH (R01CA175754, P50CA196516, P30CA142543) and CPRIT (RP130603). R.K.B. is the Michael L. Rosenberg Scholar in Medical Research and was supported by CPRIT (RP130513). Histology equipment was purchased with funding from the National Center for Advancing Translational Sciences (Center for Translational Medicine UL1TR001105).

Author Contributions W.C. designed and performed biochemical experiments; H.H., E.H., A.P.-J., Q.Y., and A.J. performed tumorigraft experiments; A.C. performed extensive statistical analyses with the supervision of X.-J.X.; M.S.K. performed RNA-seq analyses under the supervision of T.H.H. and Y.X., who also supervised H.Z. on sequencing analysis; Y.M. and N.P. performed experiments; F.H. and P.Y. performed histological analyses under the supervision of P.K.; G.H. performed PET studies under the supervision of X.S.; I.P. performed patient imaging analyses; H.G. and C.R. performed RNA-seq validation studies at the New York Genome Center; J.C. is the clinical research coordinator of the PT2385 phase 1 trial overseen by K.C. and N.Z.; K.H.G., R.K.B., and E.F. participated in discussions; T.W., J.P.R., E.M.W., and J.A.J. oversaw the development and characterization of PT2399 and provided the drug; R.M.M. assisted with manuscript preparation, writing and submission; and J.B. conceived and supervised the project, and wrote the manuscript with input from R.M.M. and the other authors.

Author Information RNAseq was released to NCBI Sequence Read Archive (SRA). ID: SRP073253. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. (james.brugarolas@utsouthwestern.edu).

METHODS

Nomenclature. Throughout the manuscript and figures, XP refers to the tumour-graft line; V refers to vehicle; S refers to sunitinib; and P refers to PT2399. Numbers following V, S, or P refer to the mouse identifier (ear tag) of that sample.

Drug trials. Drug trials in tumourgraft-bearing mice were done as previously described^{12,13}. Briefly, ~64-mm³ fragments of tissue from stably growing orthotopic tumourgrafts were implanted subcutaneously in 4–6-week-old female and male non-obese diabetic (NOD)/severe combined immunodeficiency (SCID) mice. When tumour volumes reached ~300–600 mm³, mice were segregated into treatment groups (3–5 mice per group) based on (i) tumour volume, (ii) growth rate, and (iii) mouse weight. A sample size of five mice per treatment arm gave us 80% power to detect a significant tumour volume differential on the 28th day after treatment between the reference arm and a treatment arm using a two-sample *t*-test, assuming a true 600-mm³ tumour volume difference with a standard deviation of 250 and attrition margin of ~20%. Since the mixed model analysis uses about eight repeated measures from each mouse, even with a few more covariates included in the model, the power will be similar or even higher. Vehicle (10% EtOH, 30% PEG400, 60% MCT (0.5% methyl cellulose, 0.5% Tween 80 (aq))) was administered by gavage every 12 h. Sunitinib (LC Laboratories) was administered by oral gavage every 12 h at 10 mg kg⁻¹ in 0.5% carboxymethylcellulose (CMC) in D5W. PT2399 (Peloton Therapeutics, Inc.) was administered at 100 mg kg⁻¹ by oral gavage in 10% EtOH, 30% PEG400, 60% MCT. Mouse weights were taken weekly and treatment doses were adjusted weekly. Tumours were generally measured twice a week using a digital caliper. While leading to an overestimation in tumour volumes, to minimize bias¹², tumour volume was calculated by the formula: tumour volume = $l \times w \times h$, where l is the largest dimension of the tumour, w is the largest diameter perpendicular to l , and h is maximal height of the tumour. Trials typically lasted 4 weeks, but this varied depending upon tumour growth rates. Overall, > 14,000 measurements were obtained. Assuming a digital caliper measurement error rate up to 10%, 99.8% of measurements were within protocol limits. Consideration was given to tumour growth rates, curve separation and the foreseeable need for additional mice for repeat experiments. Mice were monitored during treatment and provided appropriate veterinary care. In accordance with UT Southwestern's Institutional Animal Care and Use Committee (IACUC) policies, animals were euthanized within timeframes specified by the veterinary staff once tumour diameters were greater than 2 cm. Mice were also euthanized if they exhibited signs of adverse clinical health. A total of $n = 22$ tumourgraft trials were completed with $n = 89$ vehicle-treated tumours (sensitive: $n = 39$; intermediate: $n = 22$; resistant: $n = 28$), $n = 96$ PT2399-treated tumours (sensitive: $n = 42$; intermediate: $n = 24$; resistant: $n = 30$), and $n = 82$ sunitinib-treated tumours (sensitive: $n = 32$; intermediate: $n = 22$; resistant: $n = 28$).

Blood cell counts and haemoglobin measurements. Complete blood counts (CBC) (platelets, white blood cells, neutrophils, lymphocytes, and haemoglobin) were measured at the end of ~28-day trials and run on an IDEXX ProCyt Dx analyser. CBCs were available for 17 tumourgraft trials, with 52 vehicle-treated mice (sensitive: $n = 10$; intermediate: $n = 21$; resistant: $n = 21$), 58 PT2399-treated mice (sensitive: $n = 13$; intermediate: $n = 19$; resistant: $n = 26$), and 53 sunitinib-treated mice (sensitive: $n = 8$; intermediate: $n = 22$; resistant: $n = 23$).

PET/CT. 3'-[¹⁸F]fluoro-3'-deoxythymidine ([¹⁸F]FLT) was synthesized at 160 °C for 10 min using a GE FXN module through the nucleophilic substitution reaction between 2,3'-anhydro-5'-O-benzoyl-2'-deoxythymidine and [¹⁸F]KF (potassium fluoride) in DMSO, followed by deprotecting the benzoyl group in 1N NaOH solution. The product was separated and purified by HPLC. The injection dose of [¹⁸F]FLT was prepared in saline containing 10% ethanol. Small animal PET/CT imaging studies were performed on a Siemens Inveon PET/CT Multimodality System. PET/CT scans were conducted on mice with both orthotopic and subcutaneous tumours. Orthotopic tumourgrafts were implanted using 2–3 pieces of 2 × 2 × 2-mm tissue underneath the left renal capsule of NOD/SCID mice. Once tumours became palpable, a baseline PET/CT scan was performed, and within 72 h, PT2399 treatment was started. PT2399 was continued for 8–10 days, after which a second PET/CT scan was performed to assess tumour response. After injection with 0.12 mCi of [¹⁸F]FLT via the tail vein, and a 60-min wait period to allow for the radiotracer's distribution and uptake, mice were anaesthetized using 3% isoflurane, which was decreased to 2% during imaging. CT imaging was acquired at 80 kV and 500 μA with a focal spot of 58 μm. The PET imaging was acquired for 500 s directly following the acquisition of CT data. CT images were reconstructed with Cobra Reconstruction Software, and PET images were reconstructed using the OSEM3D algorithm. Reconstructed CT and PET images were fused and analysed using the manufacturer's software. For quantification, regions of interest were drawn aided by CT images and then quantitatively expressed as per cent injected dose per gram of tissue (%ID/g).

Immunohistochemistry. Immunohistochemistry (IHC) was performed using Dako Autostainer Link 48. The HIF-1α and HIF-2α immunohistochemical

procedures and interpretations were standardized based on expression profiles in well-characterized cell lines (786-O, 786-O empty vector, and 786-O VHL-reconstituted cell lines) and human ccRCC tissue with known expression for these two proteins by western blot. Multiple commercially available antibodies were evaluated and the antibodies with most consistent results were selected for further studies. Briefly, for HIF-1α and HIF-2α staining, after hydration, antigen retrieval was accomplished with EnVision FLEX Target Retrieval Solution, Low pH (K800521, Dako) in a Dakocytomation Pascal pressure cooker; Ki67 and CD31 antigen retrieval was done using a Dako PT Link. Slides were incubated in 3% hydrogen peroxide for 10 min. Primary antibodies were added and incubated for 40 min at room temperature. Primary antibodies: HIF-1α (1:500, NB100-105, Novus), HIF-2α (1:200, sc-46691, Santa Cruz), Ki67 (ready-to-use, IR-626, Dako) and CD31 (1:200, LS-B1932, LifeSpan BioSciences). After rinsing with wash buffer, EnVision FLEX mouse/rabbit linker (K802121/K800921, Dako) was applied to the tissue and incubated for 10 min. Secondary antibody, EnVision FLEX/HRP (K800021, Dako), was incubated for 20 min. Sections were then processed using the Envision FLEX Substrate Working Solution for 10 min followed by dehydration in a standard ethanol-xylene series and mounting medium (8310-4, Thermo Scientific). IHC of HIF-1α and HIF-2α was performed on pre-treated tumourgraft tissue for $n = 22$ tumourgraft lines. Appropriate positive and negative controls were used with each run of immunostaining. The percentage of tumour cells in the entire section examined was recorded by a pathologist blinded to the western blot results. Only a 2 or 3+ nuclear positive reaction was considered as positive expression (staining scale: 0 = no staining, 1 = weak, 2 = moderate, 3 = strong).

Proliferation index and microvessel quantification. To assess tumour proliferation index, we performed immunostaining for Ki67, and to assess tumour microvessels, we performed CD31 immunostaining on tumours following treatment with vehicle or PT2399. IHC was completed on $n = 10$ sensitive tumourgraft lines, with $n = 28$ vehicle-treated tumours and $n = 31$ PT2399-treated tumours. Slides were digitally scanned using an Aperio Scanscope AT Turbo and reviewed using the Aperio eSlide Manager (ver. 12.0.0.5040) and Imagescope (ver. 12.1.0.5029) systems (Leica Biosystems). For Ki67, Aperio Genie (ver. 11.2) pattern recognition software was used to identify and select tumour areas for quantitative analysis with the Aperio Nuclear algorithm (version 11.2), yielding a percentage of tumour nuclei positive for Ki67. In a small subset of tumours where Genie inadequately identified tumour cells, representative tumour regions were manually selected (tumour necrosis areas were avoided) and reanalysed. Quantitative measurements of microvessels, including density and average lumen area, were obtained using the Aperio Microvessel algorithm (version 11.2) from manually selected representative tumour regions.

Real-time PCR. RT-PCR data was generated for 16 tumourgraft trials, except for CA9 and LDHA, which were evaluated in 12 tumourgraft trials. Three RT-PCR reactions were run concurrently for each tumour. Total RNA was isolated as described previously³¹. cDNA was synthesized using iScript Reverse Transcription Supermix for RT-qPCR (170-8841, Bio-Rad). qRT-PCR was performed on a Bio-Rad CFX96 Real-Time PCR system using iTaq Universal SYBR Green SMX (1725124, Bio-Rad). Primers were synthesized by Invitrogen. Primers sequence available upon request.

VHL methylation. HIF2-I sensitive ccRCC tumourgrafts that had wild-type VHL status (XP164, XP373, XP453, and XP454) were tested for VHL methylation using the Affymetrix Promoter Methylation PCR Kit (MP1100).

Immunoprecipitation and western blot. Tumour tissue was lysed in IP buffer containing 25 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 1 mM EDTA, 5% glycerol, 0.5 mM DTT, with 3–4 freeze–thaw cycles. 10–20% of the lysate was saved for input; 40 μg was mixed with 3 × loading buffer (10% SDS, 33.3% glycerol, 300 mM DDT, 0.2% bromophenol blue) for input. After pre-clearing the lysate with 50 μl of a 1:1 solution of recombinant protein G-sepharose 4B (101242, Life Technologies) for 1 h, 1 mg protein was mixed with 20 μl ARNT/HIF-1β antibody (sc-55526, Santa Cruz) and rocked overnight at 4 °C. 30 μl protein G-sepharose 4B equilibrated with IP buffer was then added, rocked for 1 h at 4 °C, and spun at 3,000 rpm for 10 s. The supernatant was removed and the beads washed three times with IP buffer containing DTT. 20 μl of 1 × loading buffer was added to the beads and vortexed gently, then boiled for 5 min and spun at max speed for 5 min. The entire sample was loaded for western blot analysis. For western blot analysis, both HIF-1α antibody (A300-286A, Bethyl) and HIF-2α antibody (NB100-122, Novus) were diluted at 1:1,000 in 5% BSA in TBS and incubated overnight at 4 °C. Tubulin antibody (T5168, Sigma) was diluted at 1:5,000. Primary antibodies were detected using horseradish peroxidase-conjugated secondary antibodies (31430, 31460, Pierce) followed by exposure to enhanced chemiluminescence substrate (mixing 1:1 solution 1 (2.5 mM luminol, 0.4 mM pCoumaric acid, 0.1 M Tris-HCl) and solution 2 (0.015% H₂O₂, 0.1 M Tris-HCl)).

Sanger sequencing. Primer sequences are available upon request.

Transfections. HEK293T cells (ATCC; no perceived need for authentication; negative for mycoplasma) were cotransfected with the indicated expression plasmids using Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. After 36 h, cells were treated with PT2399 (10 μ M) at 37 °C for 5 h, harvested for immunoprecipitation with anti-FLAG beads (A2220-1ML, Sigma) and then subjected to western blot analysis. Plasmid laboratory database: #930 (pcDNA3.1 Flag-HIF1 β), #931 (pcDNA3.1 Flag-HIF1 β [F446L]), #932 (pLVX HA-HIF-2 α -IRES-zsGreen), and #933 (pLVX-HA-HIF-2 α [G323E]-IRES-zsGreen).

In silico structural analysis. The G323E and F446L mutations were evaluated using PyMOL and Protein Data Bank 4ZP4 (ref. 28).

ELISA. Mouse VEGF (MMV00), human VEGF (DVE00), and mouse erythropoietin (MEP00B) ELISA kits were from R&D Systems. Briefly, 50 μ l assay diluent was added to each well. 50 μ l of either standard, control, or sample was then added to a well. For mEPO and mVEGF ELISA, the serum was diluted twofold and fivefold, respectively, with calibrator diluent. For hVEGF, no dilution was performed. Plates were incubated for 2 h at room temperature on a horizontal orbital microplate shaker, then aspirated and washed for a total of five washes. After the last wash, 100 μ l of mEPO/mVEGF/hVEGF conjugate was added to each well and incubated for 2 h at room temperature on a shaker. Plates were washed five times with wash buffer and 100 μ l of substrate solution was added to each well and incubated for 30 min at room temperature, during which time the plates were covered to protect from the light. Stop solution was then added to each well, with gentle tapping to ensure thorough mixing. The optical density of each well was determined using a microplate reader set to 450 nm. Wavelength correction was set to 540 nm. The final optical density (OD) value was obtained by subtracting readings at 540 nm from the readings at 450 nm. ELISA data were generated for a total of 20 tumour-graft trials.

Proximity ligation assay. Mouse anti-HIF-1 α (NB100-105, Novus), mouse anti-HIF-2 α (sc-46691X, Santa Cruz) and rabbit anti-ARNT/HIF-1 β (A302-765A, Bethyl) were used. Primary antibodies were concentrated and buffers were exchanged using a Vivaspin 500 Centrifugal Concentrator (VS0131, Fisher Scientific). Antibodies were diluted to 1 mg ml⁻¹ in PBS. Primary antibody conjugation was done with a Duolink *In situ* Probenmaker MINUS/PLUS kit (DUO92010 & DUO92009, Sigma-Aldrich). Briefly, 2 μ l of conjugation buffer was added to 20 μ l of the antibody (1 mg ml⁻¹), mixed gently, transferred to one vial of lyophilized oligonucleotide (PLUS or MINUS), and incubated at room temperature overnight. 2 μ l of stop reagent was then added to the reaction and incubated at room temperature for 30 min. 24 μ l of storage solution was added and the conjugation stored at 4 °C. Tumour tissue was blocked with PBS-T (0.1% Triton X-100) + 1% BSA for 30 min after antigen retrieval. Conjugated HIF1- α -MINUS, HIF2- α -MINUS and ARNT-PLUS were diluted in blocking buffer containing 1 \times assay reagent (20 \times) at a dilution of 1:50, 1:50, and 1:200, respectively. The mixture was allowed to sit for 20 min at room temperature before diluted primary antibody was added to each sample. Slides were incubated in a humidity chamber overnight at 4 °C. Duolink *In situ* Detection Reagents Red (DUO92008, Sigma-Aldrich) were used for signal detection. Briefly, slides were washed with wash buffer A, ligation solution containing ligase at a 1:40 dilution was added, and slides were incubated in a pre-heated humidity chamber for 30 min at 37 °C. After washing in 1 \times wash buffer A with gentle agitation, amplification solution containing polymerase was added at a 1:80 dilution, and slides were then incubated in a pre-heated humidity chamber for 100 min at 37 °C. After washing in 1 \times wash buffer B and then 0.01 \times wash buffer B, slides were dried at room temperature in the dark and mounted with a cover slip using a minimal volume of Duolink *In situ* Mounting Medium with DAPI (DUO82040, Sigma-Aldrich). After approximately 15 min, slides were analysed by fluorescence microscopy (Olympus) using a 40 \times objective. Image analysis was done with the ImageJ 1.48V program. Pictures of three fields for each sample were used. At least 100 cells of each sample were counted.

RNA sequencing. 23 vehicle- and 23 PT2399-treated tumour RNA samples, including 5 sensitive XPs (XP144, XP164, XP373, XP374, and XP453) and 4 resistant XPs (XP169, XP296, XP490, and XP506), underwent RNA sequencing at the New York Genome Center. RNA sequencing libraries were prepared using the Illumina TruSeq Stranded mRNA Sample Preparation Kit. Briefly, 500 ng total RNA was purified by oligo-dT beads selecting for polyadenylated RNA species and fragmented by divalent cations under elevated temperature. The fragmented RNA underwent first strand synthesis using reverse transcriptase and random primers. Second strand synthesis created the cDNA fragments using DNA polymerase I. Following RNaseH treatment, the cDNA fragments went through end repair, adenylation of the 3' ends, and ligation of adapters. The cDNA library was enriched using eight cycles of PCR and purified. Quality control consisted of

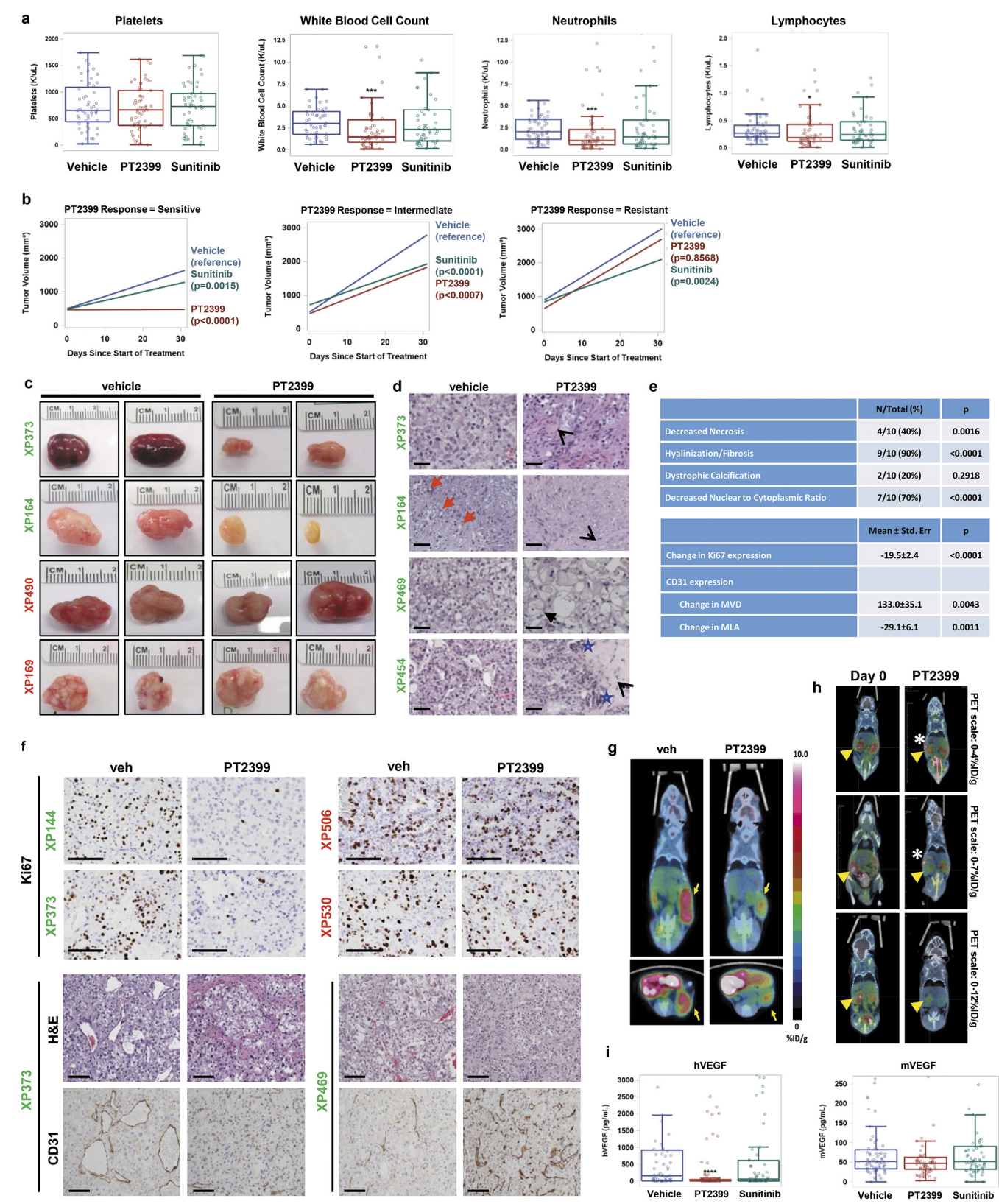
assaying the final library size using the Agilent Bioanalyzer and quantifying the final library by RT-PCR and PicoGreen (fluorescence) methods. A single peak between 250 and 350 bp indicated a properly constructed and amplified library ready for sequencing. Sequencing was performed on a HiSeq 2500 using v4 SBS chemistry according to the Illumina protocol, as described³². Sequencing libraries were loaded onto the HiSeq 2500 flowcell for clustering on the cBot using the instrument-specific clustering protocol. Given HiSeq 2500 capabilities (200–250M passed filter 2 \times 50-bp sequencing reads per flow cell lane), we sequenced 5 samples per lane in order to obtain a minimum of 50M PF reads per sample. With one exception, > 100 million reads were obtained per sample (median 146,644,355; 95% distribution-free CI: 142,380,928–151,324,826; Extended Data Table 1). Any gene with more than 50 reads in any sample was kept; only genes that had low reads in all of the samples were removed. This left 20,667 genes after removal of pseudogenes. cDNA sequences were aligned to a combined index of mouse and human reference sequences with STAR v 2.4.0c. Mouse reads were filtered out and the remaining reads were re-mapped to the NCBI hg37 using STAR aligner (v2.3.1z)³³. Quantification of genes annotated in Gencode v19 was performed using HTSeq³⁴. Picard and RSeQC³⁵ were used to collect QC metrics (<http://broadinstitute.github.io/picard/>). Differential gene expression analysis was measured using edgeR³⁶. A false discovery rate (FDR) cutoff of 0.05 was applied to identify the statistically significant genes between comparison groups. FDR was calculated using the Benjamini and Hochberg method³⁷ for adjusting for multiple hypothesis testing.

RNA-seq data were deposited into the Sequence Read Archive (SRP073253). For RNA-seq data, the tumourgraft number is preceded by an S for sensitive or R for resistant followed by treatment and ear tag. SRS1395449 (S144-P4340), SRS1395526 (S144-P4342), SRS1397028 (S164-P3281), SRS1397038 (S164-P3287), SRS1397048 (S164-P3297), SRS1397056 (S373-P4241), SRS1397057 (S373-P4244), SRS1397060 (S373-P4250), SRS1397059 (S374-P5172), SRS1397058 (S453-P5103), SRS1396986 (S453-P5104), SRS1396988 (S453-P5109), SRS1396987 (S144-V4352), SRS1396989 (S144-V4377), SRS1396991 (S164-V3290), SRS1396993 (S164-V3294), SRS1397021 (S164-V3298), SRS1397024 (S373-V4232), SRS1397025 (S373-V4236), SRS1397026 (S373-V4237), SRS1397027 (S374-V5170), SRS1397029 (S453-V5105), SRS1397031 (S453-V5107), SRS1397030 (S453-V5108), SRS1397032 (R169-P5231), SRS1397033 (R169-P5240), SRS1397034 (R169-P5241), SRS1397035 (R296-P4512), SRS1397036 (R296-P4531), SRS1397037 (R490-P3207), SRS1397039 (R490-P3210), SRS1397040 (R490-P3214), SRS1397042 (R506-P4734), SRS1397041 (R506-P4735), SRS1397043 (R506-P4736), SRS1397044 (R169-V5230), SRS1397045 (R169-V5235), SRS1397046 (R169-V5239), SRS1397047 (R296-V4519), SRS1397049 (R296-V4524), SRS1397050 (R490-V3211), SRS1397052 (R490-V3218), SRS1397051 (R490-V3224), SRS1397053 (R506-V4743), SRS1397054 (R506-V4745), SRS1397055 (R506-V4777).

Statistical analyses. Apart from the RNA-seq analysis, all reported *P* values were obtained from two-tailed tests at the 0.05 significance level. All bar charts depict the mean with the error bar representing s.e.m., while all boxplots have median centre values with fences extending to the greatest value inside the upper and lower fences (1.5[IQR] away from the 75th and 25th percentiles, respectively). Transformations were used where indicated to meet normality assumptions for analysis. These tests were completed using SAS 9.4 (SAS Institute Inc.). Except where indicated, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Regulatory. Written informed consent was obtained from the patient participating in the phase I clinical trial 'A Phase I, Dose-Escalation Trial of PT2385 Tablets In Patients With Advanced Clear Cell Renal Cell Carcinoma' (NCT02293980). UT Southwestern IACUC-approved animal protocol, APN 2015-100932, includes all live vertebrate experimental procedures.

- Peña-Llopis, S. & Brugarolas, J. Simultaneous isolation of high-quality DNA, RNA, miRNA and proteins from tissues for genomic applications. *Nat. Protocols* **8**, 2240–2255 (2013).
- Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).

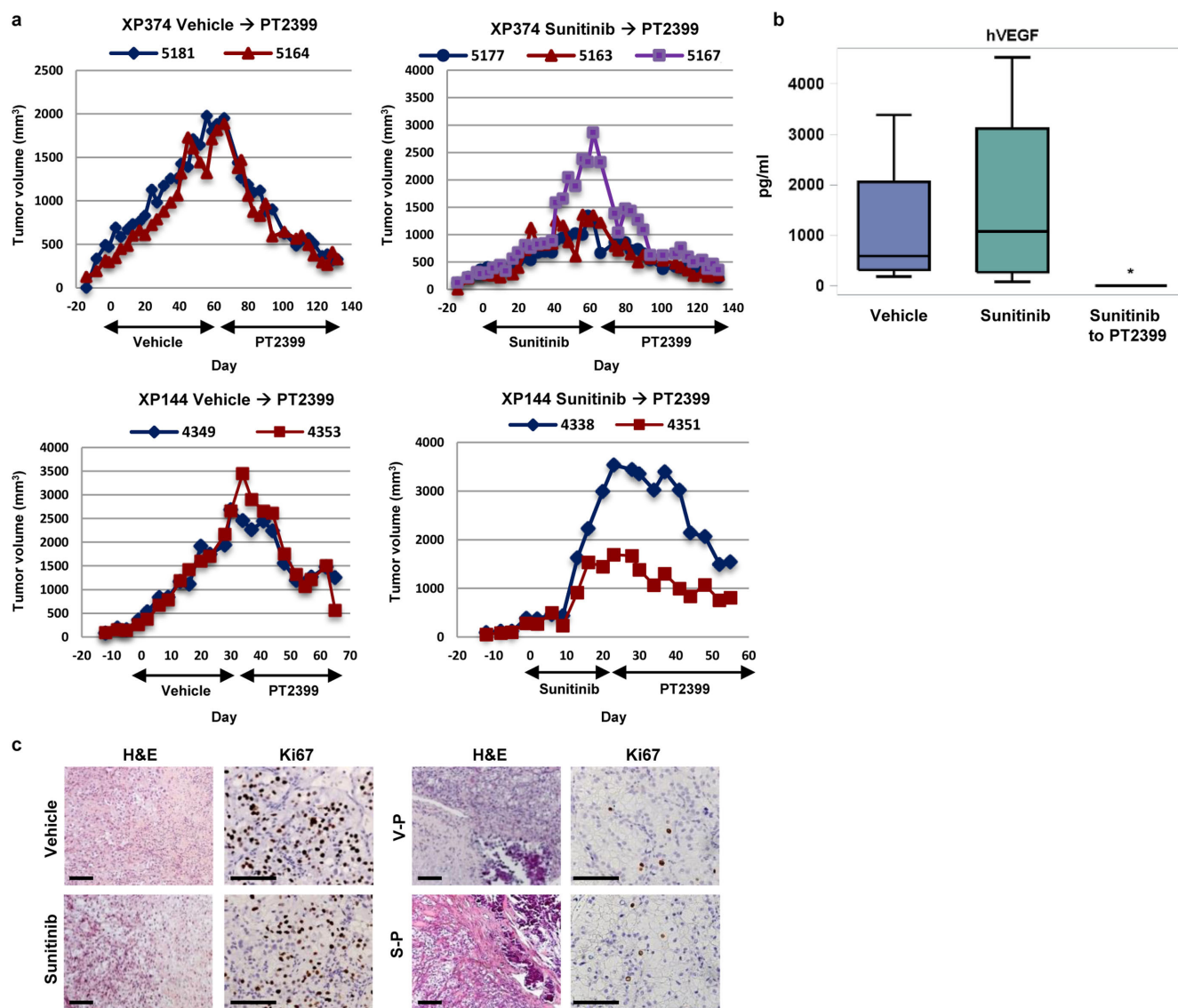


Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Effects of PT2399 on human RCC-bearing mice.

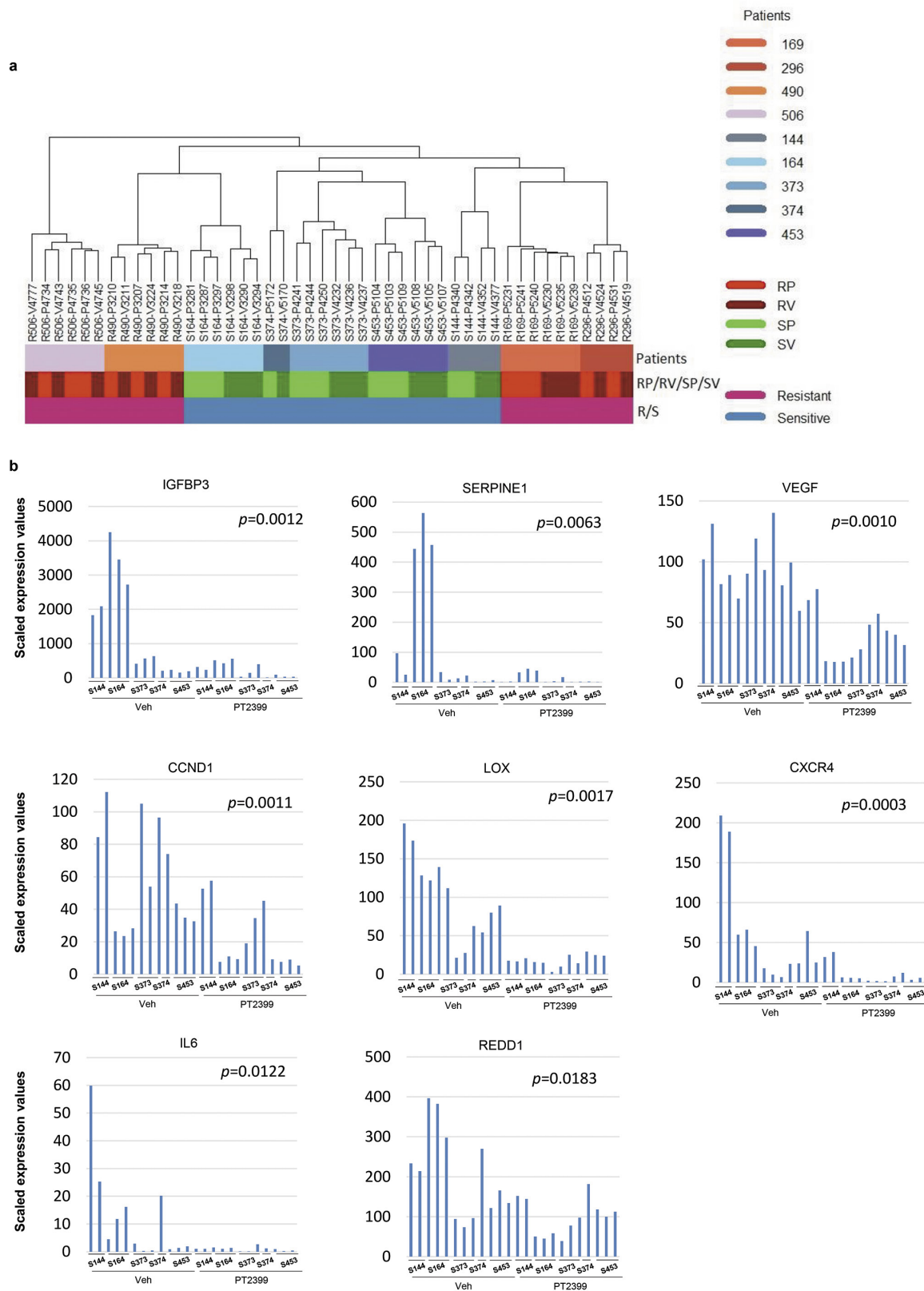
a, Platelet, white blood cell, neutrophil, and lymphocyte counts from tumourgraft-bearing mice treated with vehicle ($n = 52$), PT2399 ($n = 58$), or sunitinib ($n = 53$) at the end of the drug trial period (~28 days). Low lymphocyte levels throughout are consistent with expected levels in age- and sex-matched NOD/SCID mice. **b**, Tumour growth trend lines for sensitive, intermediate, and resistant groups after controlling for baseline tumour volume (refer to Fig. 1d for individual curves). **c**, Representative gross images of tumours from sensitive (XP164 and XP373; green) and resistant (XP169 and XP490; red) lines at the end of the drug trial. **d**, Representative haematoxylin and eosin-stained images illustrating different effects of PT2399 on sensitive tumours including patchy intercellular fibrosis and hyalinization (open arrow heads), reduced tumour necrosis (red arrows), decreased tumour cell density (XP164 and XP469), reduced nuclear-to-cytoplasmic ratio (XP469), cell ballooning (filled arrow), and dystrophic calcification (blue stars). Scale bars, 50 μm . **e**, Summary of histopathological changes induced by PT2399 in 10 sensitive tumourgraft lines represented as number of tumours (n) compared to the total or as mean \pm s.e. in 28 vehicle-treated tumours compared to 31 PT2399-treated tumours. MVD, microvessel density per mm^2 ; MLA, mean lumen area (μm^2). PT2399 collapsed tumour vasculature without decreasing the number of CD31-expressing endothelial cells. **f**, Top, IHC for Ki67 in tumours harvested from sensitive (XP144 and XP373) or resistant (XP530 and XP506) tumours following treatment with vehicle or PT2399. Bottom, haematoxylin and eosin

staining and IHC for CD31 in sensitive tumours (XP373 and XP469) treated with vehicle or PT2399. Scale bars, 100 μm . **g**, Representative [^{18}F]FLT-PET/CT images of mice with subcutaneous tumourgrafts treated with either vehicle or PT2399. Yellow arrows point to tumours. **h**, Representative [^{18}F]FLT-PET/CT images of XP144 mice with orthotopic tumours before and after treatment with PT2399 for 10 days. Yellow arrowheads, kidney tumours. White asterisks, intestine. FLT uptake in tumour compared to normal kidney reduced by 19% after 10-day treatment ($n = 3$; paired t -test, $P = 0.001$). **i**, Human and mouse VEGF levels in plasma as determined by ELISA in different treatment groups (vehicle: $n = 63$; PT2399: $n = 74$; sunitinib: $n = 61$). **a**, **i**, Tests completed using a mixed model analysis with compound symmetrical covariance structure for mice in the same tumourgraft line using vehicle as the reference group. **b**, Trend lines were obtained from a mixed model analysis for each response group using an autoregressive (1) covariance structure for the longitudinal measurements on each mouse, compound symmetry for mice within the same tumourgraft line, and controlled for baseline volume. **c**, Continuous measures were analysed using a mixed model with compound symmetrical covariance structure for mice in the same tumourgraft line and using vehicle treatment as the reference group. Specifically for categorical variables, a binomial test was used to test whether the proportion of tumours affected by PT2399 compared to vehicle was different from 10%. hVEGF and mVEGF levels were Box-Cox transformed; raw values depicted in all graphs. All boxplots have median centre values. * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$.



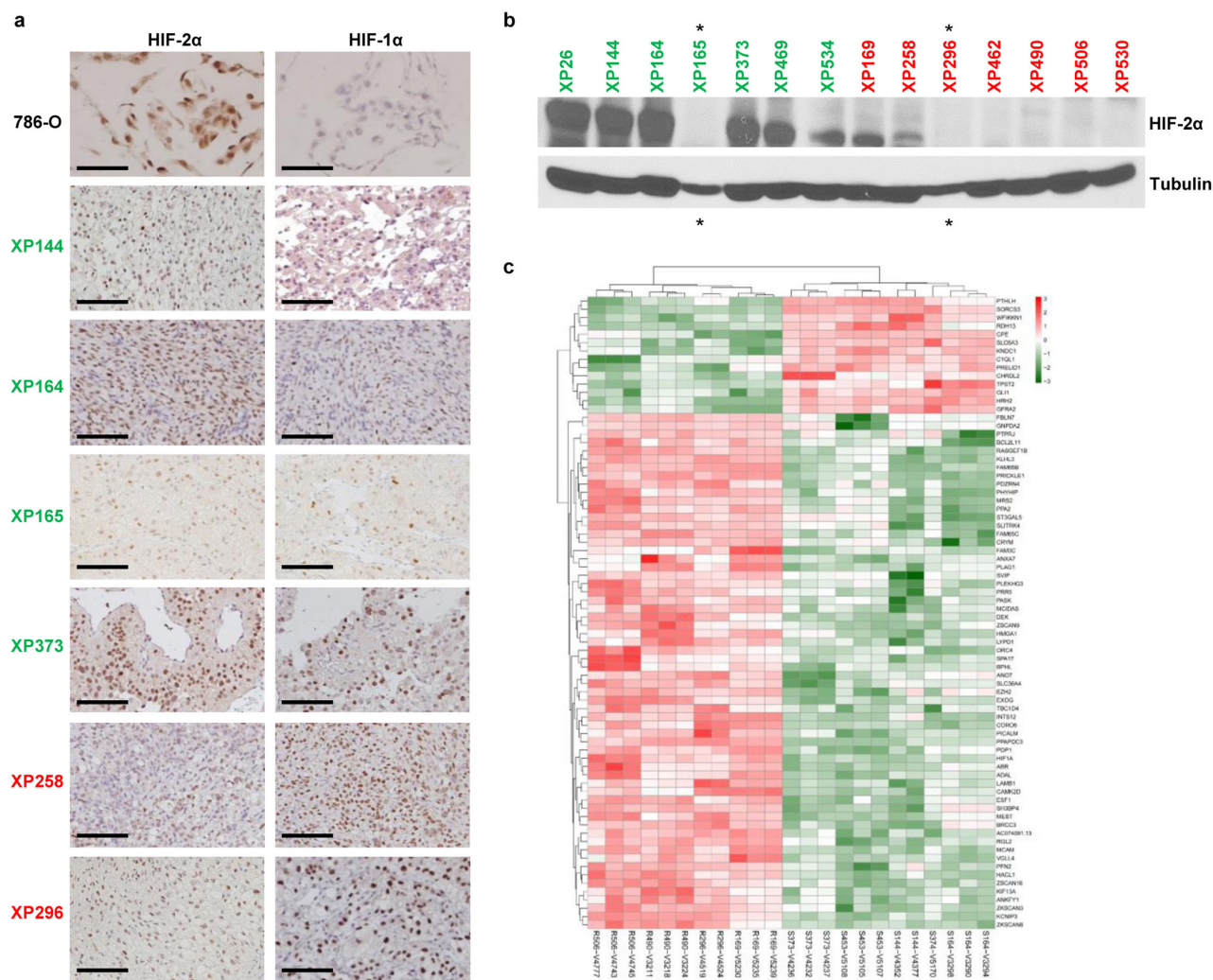
Extended Data Figure 2 | Evaluation of the effects of PT2399 on tumours progressing on sunitinib. **a**, Tumour volumes in mice from sensitive lines (XP374 or XP144) switched from vehicle or sunitinib to PT2399 as indicated (bottom black arrows). **b**, Circulating tumour-produced hVEGF levels in mice treated with vehicle, sunitinib, or sunitinib followed by PT2399. The Wilcoxon rank-sum test was used to

determine whether sunitinib ($n = 4$) or sunitinib followed by PT2399 ($n = 6$) were different from vehicle ($n = 4$). $*P < 0.05$. Boxplots have median centre values. **c**, Representative images of haematoxylin and eosin and Ki67 staining of tumours from mice (XP144) treated with vehicle or sunitinib (left) and from tumours following a switch to PT2399 (right). Scale bars, 100 μm .



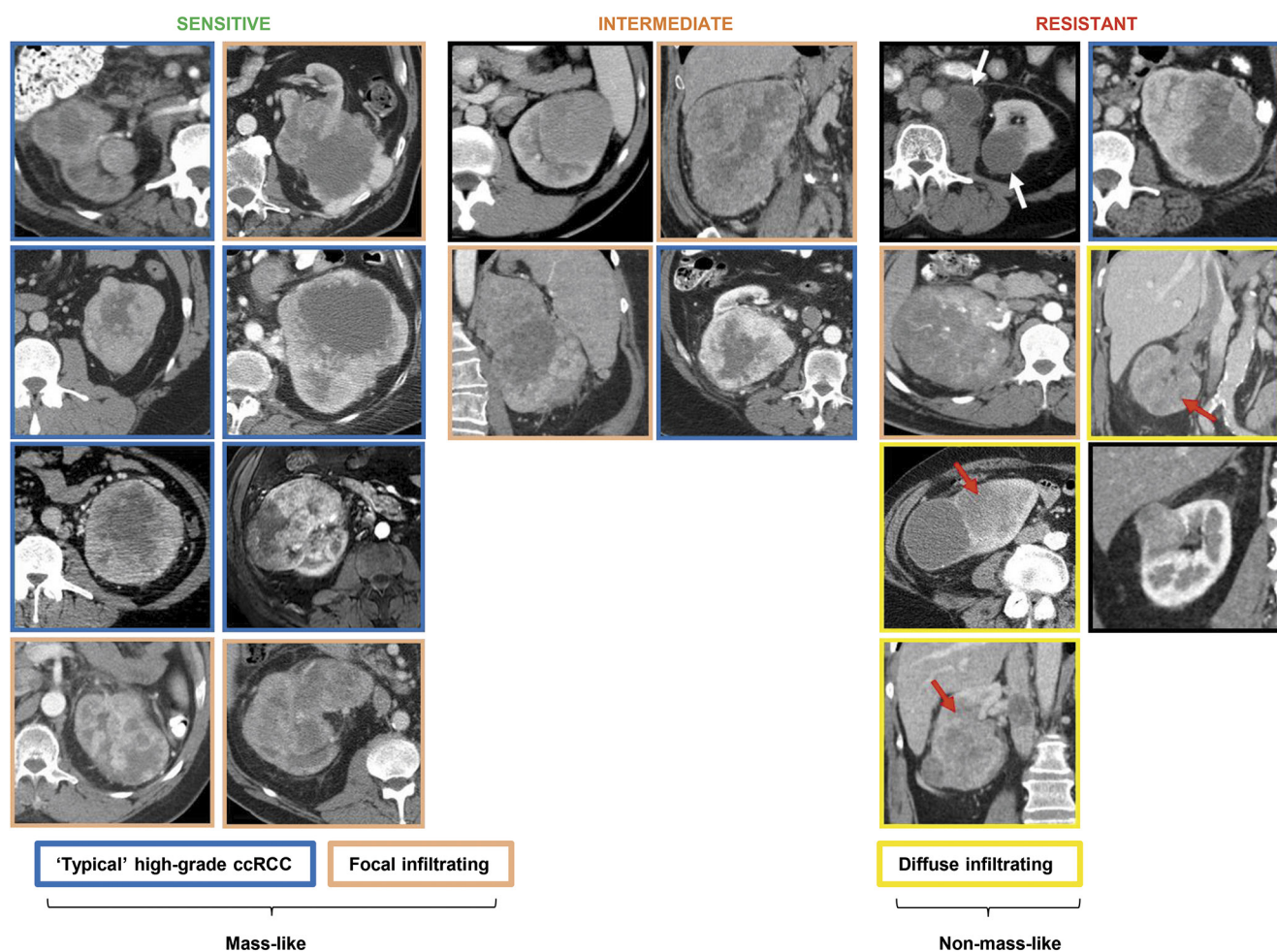
Extended Data Figure 3 | RNA-seq analyses of vehicle and PT2399-treated tumourgrafts. **a**, Unsupervised clustering analyses of all tumourgraft samples (sensitive (S) and resistant (R), both vehicle (V)- and PT2399 (P)-treated) showing clustering by tumourgraft line. **b**, RNA sequencing in sensitive tumourgrafts evaluating the effects of PT2399

on selected HIF-2 target genes. All tests completed using mixed model analysis with compound symmetrical covariance structure for mice in the same tumourgraft line. Values were \log_2 -transformed for analysis; raw values depicted in all graphs as individual bars.



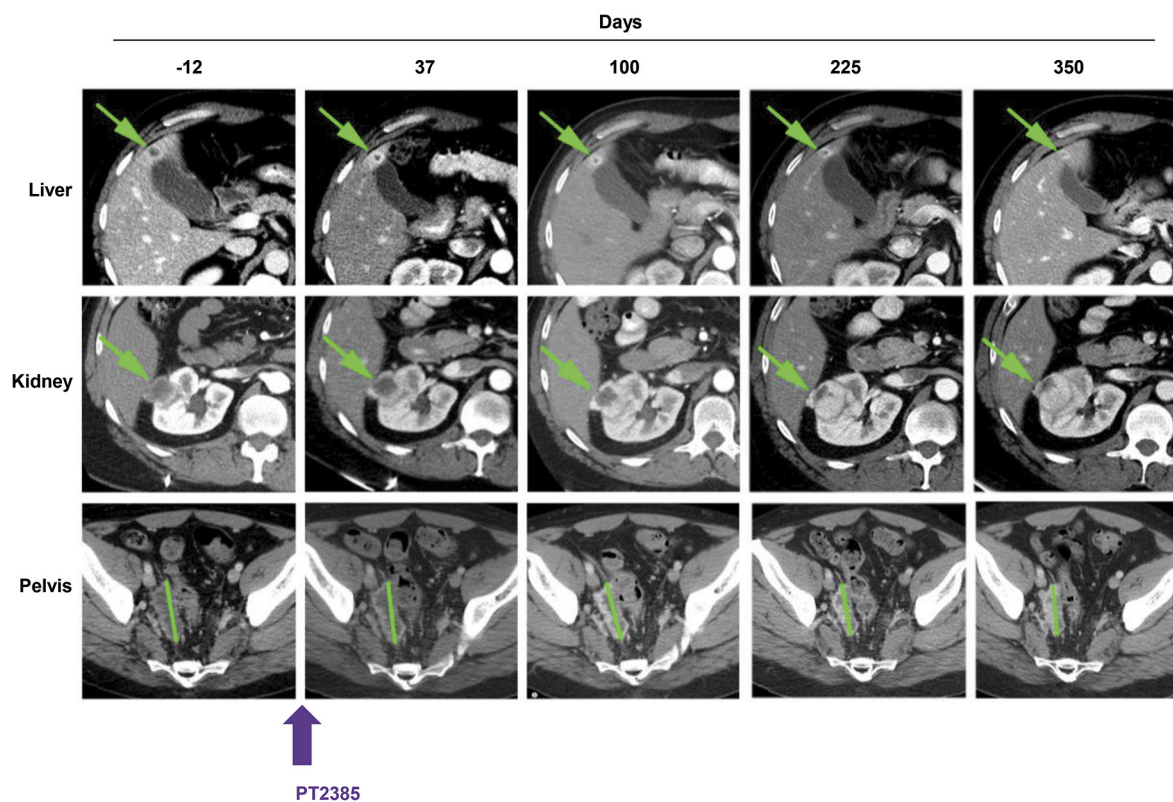
Extended Data Figure 4 | HIF-2 α and HIF-1 α levels in sensitive and resistant tumourgrafts. a, HIF-2 α and HIF-1 α IHC. 786-O cells, which express high levels of HIF-2 α , shown as controls. Scale bars, 100 μ m. **b,** Western blot analyses showing heterogeneity within tumours but with overall similar results (compare to Fig. 3c). Green, sensitive; red, resistant.

Asterisks, underloaded samples. **c,** Heatmap from RNA-seq analysis showing differentially expressed genes in sensitive (S) versus resistant (R) tumourgrafts based on uniform cutoff (see Extended Data Table 3). See Supplementary Fig. 1 for gel source images.



Extended Data Figure 5 | Evaluation of imaging characteristics of tumours in patients corresponding to sensitive, intermediate, and resistant tumourgrafts. CT scan images from patient tumours that gave rise to tumourgrafts according to tumourgraft sensitivity to PT2399. Tumours were classified into masses with peripheral hypervascularity and

a central non-enhancing area (blue outline), focally infiltrating (brown outline) and diffuse infiltrating (yellow outline). Three of the seven resistant tumours presented as non-mass-like, infiltrative neoplasms (red arrows) whereas another tumour presented with both a largely necrotic renal mass and retroperitoneal lymph nodes (black outline; white arrows).



Extended Data Figure 6 | Prolonged disease control in heavily pretreated patient with metastatic ccRCC with sensitive (XP165) tumourgraft. CT images of selected lesions in patient treated with highly related HIF-2 inhibitor (PT2385) in phase 1 clinical trial showing overall stability in the size of lesions over time. Start of treatment, day 0.

Extended Data Table 1 | Tumourgraft features

Response	XP NO.	Histology	Fuhrman Grade	Tissue	Stage at presentation	VHL status	BAP1 (IHC)	PBRM1 (IHC)	Relative GI% (p value)	RNA seq
Sensitive	XP26	ccRCC	2	Adrenal	pT1aNxMx	mut	mut	wt	87 (0.0003)	
	XP144	ccRCC	4	Kidney	pT2aN1Mx	mut	wt	wt	98 (<0.0001)	Y
	XP164	ccRCC*	4	Kidney	pT4NxMx	wt [‡]	wt	wt	134 (<0.0001)	Y
	XP165	ccRCC	3	Abd wall	pT3bNxMx	mut	wt	mut	112 (<0.0001)	
	XP373	ccRCC	4	Tu Thr	pT3aN1M1	wt [‡]	mut	wt	103 (<0.0001)	Y
	XP374	ccRCC*#	4	Kidney	pT4NxMx	mut	wt	wt	109 (<0.0001)	Y
	XP453	ccRCC	3	Tu Thr	pT3bN0Mx	wt [‡]	wt	wt	110 (<0.0001)	Y
	XP454	ccRCC	3	Kidney	pT3aN0Mx	wt [‡]	mut	wt	156 (0.0032)	
	XP469	ccRCC	3	Kidney	pT3bN0Mx	mut	mut	mut	91 (<0.0001)	
	XP534	ccRCC	4	Kidney	pT3aN0Mx	mut	mut	n/a	129 (0.0007)	
	XP237	tRCC	n/a	LN	pT3aN1M1	n/a	wt	wt	43 (0.0144)	
	XP391	ccRCC	4	Tu Thr	pT3bN0Mx	mut	wt	mut	45 (0.0018)	
Intermediate	XP426 [@]	ccRCC#	4	Kidney	pT4N1Mx	mut	wt	wt	44 (0.0273)	
	XP427 [@]	ccRCC#	4	LN	pT4N1Mx	mut	wt	wt	54 (0.0206)	
	XP466	ccRCC	3	Kidney	pT3aN0M1	wt	mut	n/a	67 (0.0030)	
	XP169	Unclassified	n/a	Kidney	pT4N1M1	wt	wt	wt	0 (0.0119 [§])	Y
	XP258	ccRCC*	4	Kidney	pT3aN0M1	mut	mut	wt	39 (0.11)	
	XP296	ccRCC*	4	Kidney	pT3aNxM1	mut	wt	wt	29 (0.30)	Y
	XP462	Unclassified	n/a	Kidney	pT3aN0M1	wt	wt	mut	29 (0.11)	
	XP490	ccRCC*#	4	Kidney	pT3aN1M1	mut	wt	wt	39 (0.89)	Y
	XP506	ccRCC	3	Ascites	pT3aN1Mx	wt [‡]	wt	wt	20 (0.76)	Y
	XP530	Unclassified	n/a	Kidney	pT3bN0Mx	wt	n/a	n/a	2 (0.68)	
Resistant										

Fuhrman grade of primary tumour and stage at presentation (metachronous metastasis may have developed). Tissue, engrafted tissue; IHC, immunohistochemistry from tumourgrafts; GI, growth inhibition; ccRCC, clear-cell renal cell carcinoma; tRCC, translocation renal cell carcinoma; Abd, abdominal; Tu Thr, tumour thrombus; LN, lymph node; mut, mutant; wt, wild-type.

[@]Independent tumours from same patient.

*Sarcomatoid differentiation.

#Rhabdoid features.

[§]PT2399-treated mice had greater relative growth than vehicle-treated mice.

[‡]Promoter methylation.

[‡]Promoter not methylated.

Extended Data Table 2 | RNA sequencing read data

Samples	Read Count	Samples	Read Count
S144-P4340	131,078,351	R169-P5231	150,980,881
S144-P4342	127,945,953	R169-P5240	146,751,739
S164-P3281	121,045,606	R169-P5241	144,959,159
S164-P3287	128,070,443	R296-P4512	151,324,826
S164-P3297	138,586,535	R296-P4531	144,982,512
S373-P4241	162,092,320	R490-P3207	142,380,928
S373-P4244	146,116,441	R490-P3210	164,412,241
S373-P4250	140,629,410	R490-P3214	169,970,555
S374-P5172	88,374,928	R506-P4734	165,472,466
S453-P5103	120,921,569	R506-P4735	154,474,148
S453-P5104	108,148,316	R506-P4736	173,988,590
S453-P5109	117,009,388	R169-V5230	159,863,685
S144-V4352	128,119,810	R169-V5235	146,783,488
S144-V4377	148,456,002	R169-V5239	144,377,378
S164-V3290	144,464,174	R296-V4519	146,536,970
S164-V3294	161,750,684	R296-V4524	148,798,769
S164-V3298	152,823,172	R490-V3211	162,273,604
S373-V4232	156,310,574	R490-V3218	123,559,977
S373-V4236	150,155,973	R490-V3224	151,672,989
S373-V4237	148,496,505	R506-V4743	181,173,536
S374-V5170	130,903,402	R506-V4745	156,598,756
S453-V5105	123,966,544	R506-V4777	164,427,358
S453-V5107	123,347,998		
S453-V5108	112,341,672		

Samples are labelled as S (sensitive) or R (resistant) followed by the tumourgraft (XP) line, treatment type (P, PT2399 or V, vehicle) and mouse identifier (ear tag).

Extended Data Table 3 | Number of differentially regulated RNAs across tumourgraft groups by RNA-seq analysis

(Cutoff)	SP vs. SV		RV vs. SV		RP vs. SP	
	Up	Down	Up	Down	Up	Down
Original (FDR < 0.05)	195	297	1776	1766	1640	1815
T-Test (P-value < 0.01)	99	213	852	584	798	695
Wilcox (P-value < 0.01)	90	207	829	498	760	621
Uniform (all $\Delta > 0$ or $\Delta < 0$)	2	5	78	16	61	15

S, sensitive; R, resistant; V, vehicle; P, PT2399; FDR, false discovery rate.

Extended Data Table 4 | Top 15 downregulated and top 15 upregulated pathways in sensitive tumours treated with PT2399

Ingenuity Canonical Pathways	Molecules	P-Value
Top 15 Downregulated pathways		
Cell Cycle Control of Chromosomal Replication	CDC7,ORC1,MCM7,CDC45,MCM2,CDK2,ORC6,MCM3,MCM6,MCM4,MCM5,CDC6	7.94328E-16
Mitotic Roles of Polo-Like Kinase	KIF23,ESPL1,PRC1,CDC20,CCNB2,PPP2R2A,CDC7,PLK4,PTTG1,PLK1,CDK1,CCNB1,FBXO5	3.98107E-12
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	TOP2A,BRCA1,AURKA,CKS1B,CKS2,CCNB2,PLK1,CDK1,CCNB1	2.51189E-08
Cyclins and Cell Cycle Regulation	CCNE2,SUV39H1,CCND1,CDK2,E2F1,CCNA2,CCNB2,PPP2R2A,CDK1,CCNB1	1.31826E-07
ATM Signaling	JUN,BLM,BRCA1,FANCD2,CDK2,RAD51,CCNB2,CDK1,CCNB1	1.38038E-07
GADD45 Signaling	CCNE2,BRCA1,CCND1,CDK2,CDK1,CCNB1	1.77828E-07
DNA damage-induced 14-3-3 σ Signaling	CCNE2,BRCA1,CDK2,CCNB2,CDK1,CCNB1	1.77828E-07
Estrogen-mediated S-phase Entry	CCNE2,CCND1,CDK2,E2F1,CCNA2,CDK1	8.31764E-07
Role of BRCA1 in DNA Damage Response	BLM,BRCA1,FANCB,FANCD2,E2F1,FANCA,BRCA2,RAD51,PLK1	1.58489E-06
Hereditary Breast Cancer Signaling	BLM,BRCA1,FANCB,FANCD2,CCND1,E2F1,FANCA,BRCA2,RAD51,CDK1,CCNB1	1.8197E-06
Role of CHK Proteins in Cell Cycle Checkpoint Control	BRCA1,CDK2,E2F1,CLSPN,PPP2R2A,PLK1,CDK1	1.20226E-05
Ovarian Cancer Signaling	BRCA1,VEGFA,SUV39H1,LEF1,PTGS2,CCND1,E2F1,PRKAG2,BRCA2,RAD51	1.65959E-05
Pancreatic Adenocarcinoma Signaling	VEGFA,SUV39H1,PTGS2,CCND1,BIRC5,CDK2,E2F1,BRCA2,RAD51	1.99526E-05
Aryl Hydrocarbon Receptor Signaling	JUN,CCNE2,MCM7,CCND1,IL1A,CDK2,E2F1,CCNA2,IL6,ALDH1A3	2.29087E-05
Small Cell Lung Cancer Signaling	CCNE2,SUV39H1,PTGS2,CCND1,CDK2,E2F1,CKS1B	6.45654E-05
Top 15 Upregulated pathways		
Hepatic Fibrosis / Hepatic Stellate Cell Activation	COL19A1,COL4A6,FGFR2,IGFBP4,TGFB3,MYL4,PDGFD,COL21A1,MYL3,IGF1	8.13E-06
Axonal Guidance Signaling	GLIS2,CXCL12,BMP5,MYL4,PDGFD,EPHB3,SRGAP3,SMO,SEMA3F,IGF1,ADAMTS1,SEMA3G,BMP4,MYL3	2.09E-05
Human Embryonic Stem Cell Pluripotency	FGFR2,TGFB3,BMP5,PDGFD,BMP4,LEFTY2,SMO	0.000145
PAK Signaling	MYL4,PDGFD,EPHB3,MYLK,MYL3	0.001
Basal Cell Carcinoma Signaling	GLIS2,BMP5,BMP4,SMO	0.00302
RhoA Signaling	MYL4,MYLK,SEMA3F,MYL3,IGF1	0.00389
Agranulocyte Adhesion and Diapedesis	CXCL12,MYL4,MYL3,CCL11,CXCL14,MMP24	0.004266
Cellular Effects of Sildenafil (Viagra)	MYL4,ADCY5,MYLK,MYL3,PDE5A	0.004467
TR/RXR Activation	THRSP,HIF1A,G6PC,AKR1C1/AKR1C2	0.00631
Regulation of Actin-based Motility by Rho	DIRAS3,MYL4,MYLK,MYL3	0.006918
Factors Promoting Cardiogenesis in Vertebrates	TGFB3,BMP5,BMP4,SMO	0.007413
CXCR4 Signaling	DIRAS3,CXCL12,MYL4,ADCY5,MYL3	0.010233
Cardiomyocyte Differentiation via BMP Receptors	BMP5,BMP4	0.011482
Cardiac Hypertrophy Signaling	DIRAS3,TGFB3,MYL4,ADCY5,MYL3,IGF1	0.012303
Molecular Mechanisms of Cancer	DIRAS3,TGFB3,BMP5,BCL2L11,BMP4,ADCY5,HIF1A,SMO	0.013183

Atomic model for the membrane-embedded V_O motor of a eukaryotic V-ATPase

Mohammad T. Mazhab-Jafari¹, Alexis Rohou², Carla Schmidt^{3†}, Stephanie A. Bueler¹, Samir Benlekibir¹, Carol V. Robinson³ & John L. Rubinstein^{1,4,5}

Vacuolar-type ATPases (V-ATPases) are ATP-powered proton pumps involved in processes such as endocytosis, lysosomal degradation, secondary transport, TOR signalling, and osteoclast and kidney function. ATP hydrolysis in the soluble catalytic V_1 region drives proton translocation through the membrane-embedded V_O region via rotation of a rotor subcomplex. Variability in the structure of the intact enzyme has prevented construction of an atomic model for the membrane-embedded motor of any rotary ATPase^{1–5}. We induced dissociation and auto-inhibition of the V_1 and V_O regions of the V-ATPase by starving the yeast *Saccharomyces cerevisiae*^{6,7}, allowing us to obtain a ~ 3.9 -Å resolution electron cryomicroscopy map of the V_O complex and build atomic models for the majority of its subunits. The analysis reveals the structures of subunits $c_8c'c''$ and a protein that we identify and propose to be a new subunit (subunit f). A large cavity between subunit a and the c-ring creates a cytoplasmic half-channel for protons. The c-ring has an asymmetric distribution of proton-carrying Glu residues, with the Glu residue of subunit c'' interacting with Arg735 of subunit a. The structure suggests sequential protonation and deprotonation of the c-ring, with ATP-hydrolysis-driven rotation causing protonation of a Glu residue at the cytoplasmic half-channel and subsequent deprotonation of a Glu residue at a luminal half-channel.

Following starvation-induced dissociation of the V_1 and V_O regions of V-ATPase, the auto-inhibited V_O complex is markedly more structurally homogeneous than in the intact V-ATPase¹. Our 3D electron cryomicroscopy (cryo-EM) map reveals the structure of the intact V_O complex (Fig. 1a–c, Extended Data Fig. 1 and Supplementary Table 1) with an overall resolution of 3.9 Å (Extended Data Fig. 2). Although the resolution is variable throughout the map (Extended Data Fig. 2c), side-chain densities are apparent for most of the α -helices in the complex, allowing for the construction of an atomic model (Fig. 1d and Extended Data Fig. 3). A few regions (for example, subunits $c_{(5)}$ and $c_{(6)}$ in Extended Data Fig. 3g) only have sufficient resolution to model large side chains; however this still allows the protein sequence to be placed in register in the map. Resolution was poorest for the soluble domain of subunit a where it contacts subunit d, which we modelled entirely as poly-alanine. The map reveals the structures of the $c_8c'c''$ -ring (Fig. 1a, pink, magenta and purple), subunit a (Fig. 1a, green), subunit d (Fig. 1a, cyan), and two additional proteins. These additional components were detected in a lower-resolution analysis of the intact V-ATPase complex^{1,5}, but can be seen here to consist of two transmembrane α -helices each (Fig. 1b, c, blue and red–brown). Both of these proteins appear to be stoichiometric components of the complex, as suggested by their relative densities in the map. One of these subunits (Fig. 1b, c, red–brown) abuts α -helix 2 and the loop connecting α -helices 7 and 8 of subunit a. The other protein is an α -helical hairpin that contacts α -helix 1 of subunit a (Fig. 1b, c, blue). These two proteins are discussed in more detail later on.

Subunit d sits on top of the c-ring, linking the V_1 and V_O parts of the central rotor. As seen previously in an ~ 18 -Å resolution map of the V_O complex⁸, subunit d sits deeper within the c-ring in the V_O complex than it does in the intact V-ATPase¹. In contrast to the intact V-ATPase structure, in the V_O complex the N-terminal domain of subunit a from the stator part of the enzyme is in contact with subunit d from the rotor^{1,8,9}. This interaction would prevent rotation of the c-ring, consistent with the observation that the V_O complex is impermeable to protons when separated from the V_1 region^{6,7}. However, the V_O complex remains impermeable to protons even when subunit d or the N-terminal domain of subunit a are removed^{8,10}. Furthermore, we also identified a 3D class that lacks subunit d but otherwise appears to have almost the same structure as the intact V_O complex (Extended Data Fig. 4). Consequently, the movement of the N-terminal domain of subunit a, and inhibition of proton translocation, cannot be due to interaction of the N-terminal domain of subunit a with subunit d.

The fold of subunit a is consistent with the fold predicted by lower-resolution cryo-EM studies of the intact V-ATPase combined with evolutionary covariance analysis⁵. The subunit contains eight membrane-embedded α -helices, with structured loops connecting some of these α -helices. The membrane-embedded domain of subunit a starts with a pair of short α -helices that do not fully cross the lipid bilayer (Fig. 1b, c, α -helices a_1 and a_2). Four subsequent transmembrane α -helices (a_3 to a_6) produce a central layer in the subunit structure, which terminates with two long and highly-tilted transmembrane α -helices that contact the c-ring (a_7 and a_8) and are characteristic of rotary ATPases^{1–3,5}. No density was detected for the loop between residues 659 and 709, which is the region with the least sequence similarity amongst the different isoforms of subunit a⁵. The lack of density for this loop suggests that it is mobile in the structure. Another loop (residues 481–523) could also not be modelled, although a region of low-resolution density in the map (Fig. 1b, semi-transparent grey density) probably corresponds to this loop and to parts of the two small membrane-embedded proteins in the V_O complex.

Unlike the c-ring of F-type ATP synthases and bacterial V/A-ATPases, the proton-carrying ring of the eukaryotic V-ATPase is hetero-oligomeric. The *S. cerevisiae* V-ATPase $c_8c'c''$ -ring has 40 transmembrane α -helices arranged into an inner ring and outer ring of 20 α -helices each^{1,11} (Fig. 2). Each c, c' , and c'' subunit contributes four α -helices to the c-ring, two to the inner ring and two to the outer ring. The cryo-EM map shows two additional α -helices that pass through the $c_8c'c''$ -ring (Fig. 2b, purple asterisk). These α -helices can be attributed to 56 residues at the N terminus of subunit c'' (ref. 12), most of which are not necessary for proton translocation¹³ and correspond to density previously seen within the c-ring¹⁴ and upon re-examination of the intact V-ATPase maps¹. The first additional α -helix is adjacent to subunit $c_{(1)}$ and the second is near the centre of the c-ring with a clear connection to the rest of subunit c'' . Although map resolution is limited

¹Molecular Structure and Function Program, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. ²Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. ³Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK. ⁴Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 1L7, Canada. ⁵Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [†]Present address: Interdisciplinary Research Center HALOmern, Martin Luther University, Halle-Wittenberg, 06120 Halle (Saale), Germany.

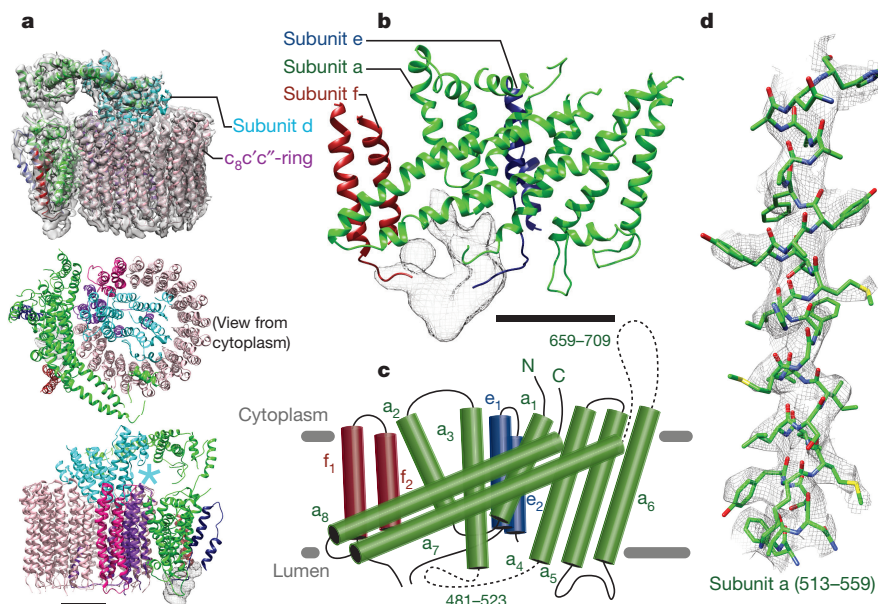


Figure 1 | The intact V_0 complex. **a**, The V_0 cryo-EM density map (semi-transparent surface, top) shows subunits a (green), c (pink), c' (magenta), c'' (purple), d (cyan), e (blue), and f (red-brown). The asterisk indicates the contact between subunits d and c'' . **b**, **c**, An enlarged view and cartoon of

the C-terminal domain of subunit a with subunits e and f. Unfitted density is shown as grey mesh. **d**, An illustration of how the cryo-EM map allows atomic models to be built for most of the protein sequence. Scale bars, 25 Å.

in this region, it is sufficient to register the sequence of the α -helices (Extended Data Fig. 3b). This structure for subunit c'' differs from a crystal structure of a bacterial hybrid F/V-ATPase c-ring, in which a single additional N-terminal helix lay across the periplasmic surface of the ring¹⁵. Subunit c'' serves as the main contact between subunit d and the c-ring (Fig. 1a, bottom, cyan asterisk), making it critical for the transmission of ATP-driven rotation of subunits D and F of the V_1 region to the c-ring of the V_0 region. Comparison to the lower-resolution cryo-EM maps of the intact V-ATPase¹ (Extended Data Fig. 5a–c) shows that the V_0 complex is in the least populated, and probably the least stable, of the three rotational states identified for the intact enzyme (Extended Data Fig. 5d). Therefore, subunit c'' apparently marks the position of the c-ring from which the complex can disassemble and reassemble^{6,7}. An unexpected feature of the subunit c'' structure is the location of its essential proton-carrying Glu residue¹². Each c, c' , and c'' subunit has a single proton-carrying Glu residue (Glu137, 145, and 108, respectively¹²), giving 10 proton-carrying residues for the 20 outer α -helices of the ring. Notably, the Glu residue of subunit c'' is on its second ring-forming α -helix, whereas in subunits c and c' it is found on the fourth ring-forming α -helix. This arrangement gives an irregular distribution of acidic proton-carrying Glu residues, rather than having proton-binding sites on alternating α -helices (Fig. 2b, right, in red). The two adjacent negative charges of subunits c'' and $c_{(1)}$ that are in contact with subunit a may determine the disassembly- and reassembly-competent conformations of the enzyme.

Transmembrane proton pumping is thought to occur via two offset half-channels: a cytoplasmic half-channel that protonates Glu residues of the c-ring and a luminal half-channel that deprotonates Glu residues^{16,17}. The cytoplasmic half-channel, which leads to Glu108 of subunit c'' , is readily apparent from inspection of the model as a deep cavity between the c-ring and the two highly tilted α -helices of subunit a (Fig. 3a, b, blue density and arrow). At its opening to the cytoplasm, this cavity is almost 20 Å across and 10 Å wide and is predicted to be filled with water¹⁸, providing direct access to Glu108 for protons from the cytoplasm. The location of this channel, at the interface of subunit a and the c-ring, is consistent with biochemical experiments^{19,20} and with the indentations observed in the detergent micelles of lower-resolution cryo-EM density maps of the F-, V/A-, and V-type ATPases^{1–3,5}. The position of the cytoplasmic half-channel

places it in close proximity to the loop of 51 residues between α -helices a_5 and a_6 (659–709), which contains 21 charged residues⁵. This loop may be involved in modulating access to the half-channel⁵. The location of the luminal half-channel is less obvious. Inspection of lower-resolution cryo-EM maps suggested the presence of a gap in the detergent micelle²¹. Consideration of residues conserved across species suggested a proton path that involves the first and second membrane-embedded α -helices of subunit a⁵ (Fig. 1c, a_2 and a_3). If this position is the presumed location of the luminal half-channel, its exit is thus located near to the region of unmodelled density in the cryo-EM map (Fig. 3a, grey mesh). However, a clear path through the protein, from the centre of the lipid bilayer to its luminal surface, is not readily

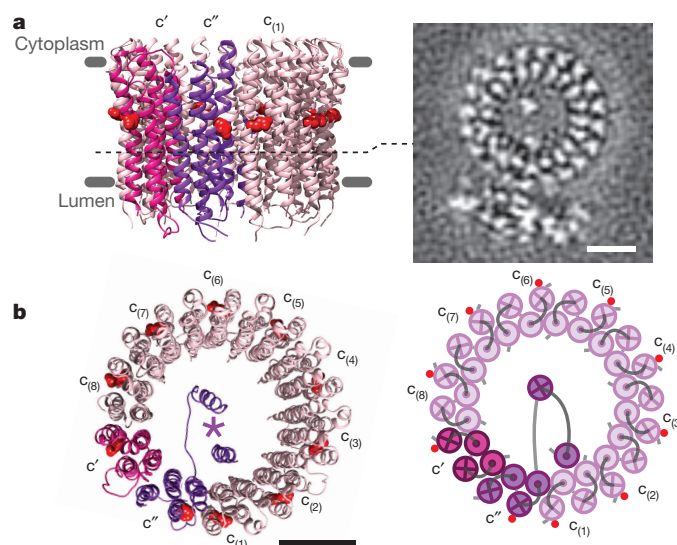


Figure 2 | Asymmetry of the c-ring. **a**, A ribbon model of the c-ring (left) with a map section (right) shows the two N-terminal α -helices of subunit c'' in the middle of the c-ring. **b**, A view of the c-ring model (left) and cartoon (right) from the cytoplasm shows the asymmetric distribution of mid-membrane Glu residues around the ring (red). The two N-terminal α -helices of subunit c'' are marked with an asterisk (left). Scale bars, 25 Å.

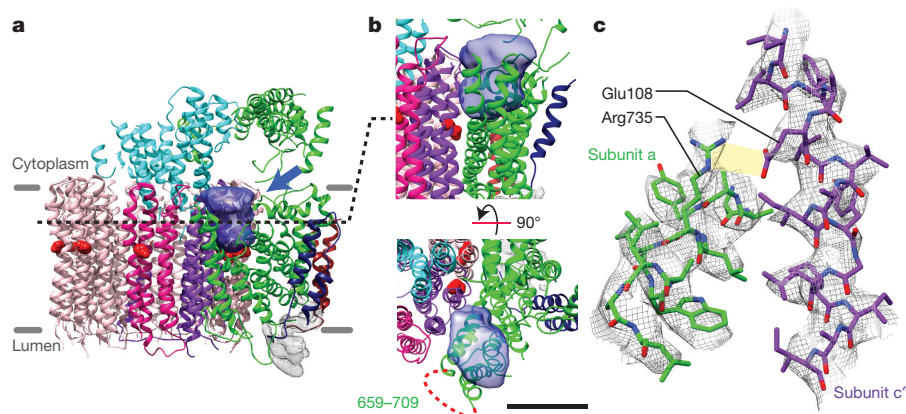


Figure 3 | Cytoplasmic half-channel and subunit a/c-ring interaction. **a, b,** A large cavity (blue density and arrow) is apparent between the c-ring and subunit a in the expected position of the cytoplasmic half-channel. The cavity is near the unresolved loop composed of residues 659 to 709

apparent from the model and could not be detected through automated identification of potential pore-lining residues²² or by searching for gaps in the protein²³. It should be recalled that the dissociated V_O region is auto-inhibited and does not allow for proton translocation^{6,8} and therefore the luminal half-channel may not exist in this state. Alternatively, the luminal half-channel may remain closed in proton-pumping V-ATPases, only opening transiently during proton translocation. Future investigation of the path that protons take from the centre of the lipid bilayer to the lumen is therefore warranted.

from subunit a (bottom of **b**, dashed red line). Scale bar, 25 Å. **c,** At the mid-membrane terminus of the cavity, the essential residues Arg735 from subunit a and Glu108 from subunit c'' are positioned to interact.

All rotary ATPase a subunits include an essential Arg residue (Arg735 in the *S. cerevisiae* V-ATPase) that is necessary to couple ATP synthesis or hydrolysis to proton translocation^{24,25}. The conserved Glu residues of subunits c, c', and c'' are also essential for proton transport in the V-ATPase¹². In the auto-inhibited V_O complex, Glu108 of subunit c'' interacts with Arg735 present on the penultimate α -helix of subunit a⁵ (Fig. 3c). This interaction suggests that Glu108 is deprotonated. The presumed salt bridge between Arg735 and Glu108 is similar to those between Arg–Asp pairs shown to ensure proton selectivity in

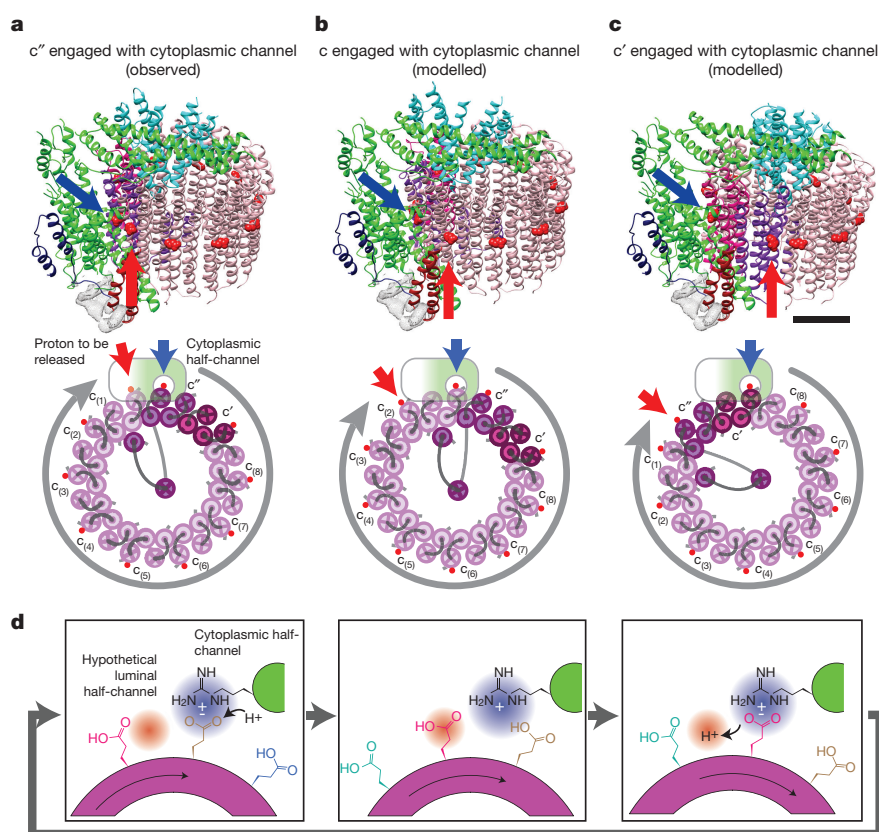


Figure 4 | Irregularly spaced Glu residues require protonation before deprotonation. **a,** The observed structure (top) with cartoon (bottom) illustrates that when Glu108 of subunit c'' is aligned with the cytoplasmic half-channel (blue arrow), Glu137 of subunit c₍₁₎ contacts subunit a (red arrow) and could be aligned with a luminal half-channel. **b,** When Glu137 of a subunit c is aligned with the cytoplasmic half-channel (blue arrow)

Glu137 of the next subunit c is not in contact with subunit a (red arrow). **c,** When Glu145 of subunit c' is aligned to the cytoplasmic half-channel (blue arrow) Glu108 of subunit c'' is too far from subunit a to interact (red arrow). Scale bar, 25 Å. **d,** The sequence of protonation and deprotonation events that occur during c-ring rotation, with each Glu residue given its own colour.

voltage-gated proton channels²⁶. However, because rotary ATPases can transport different ions and the architecture of rotary ATPases appears to be generally conserved, the function of the Arg–Glu pair here is probably to ensure that the Glu residue is stripped of its ion before it leaves the luminal half-channel, rather than to ensure proton specificity.

Liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis of in-gel-digested protein suggested that Vma9p (subunit e) and putative protein YPR170W-B were present in the purified V_O complex (Supplementary Table 2 and Supplementary Data). The α -helical hairpin shown in blue in Fig. 1a–c was identified as subunit e because its density fit with that of the three Trp residues and the Phe and Tyr residues of subunit e (Extended Data Fig. 3e). The structure of subunit e, and the interaction of its C terminus with a loop in subunit a, explains why addition of a C-terminal affinity tag to subunit e causes its dissociation from the detergent-solubilized complex²⁷. Its position distal to the area of interaction between subunit a and the c-ring is consistent with the observation that its removal from the assembled V-ATPase does not affect proton-pumping activity²⁷. Deletion of the *VMA9* gene leads to the conditionally lethal VMA phenotype²⁸ and the absence of V-ATPase from the yeast vacuole, suggesting that subunit e is necessary for the successful localization and assembly of the complex²⁹. To confirm the presence of putative protein YPR170W-B as the additional membrane-embedded component of the V_O complex, we constructed a *S. cerevisiae* strain with a 3 \times FLAG-tag fused to the C terminus of YPR170W-B. Affinity purification from the detergent-solubilized membranes of this strain isolated the V_O complex with some subunits from the intact V-ATPase (Extended Data Fig. 6a), confirming the protein as a component of the V_O complex. We subsequently purified the V_O complex from a strain of yeast lacking YPR170W-B and determined its structure by cryo-EM (Extended Data Fig. 6b). The missing density in the 3D map allowed us to unambiguously locate YPR170W-B in the complex. Deletion of YPR170W-B, which we now tentatively identify as subunit f of the V-ATPase, did not produce the VMA phenotype²⁸ (Extended Data Fig. 6c), indicating that it is not essential for V-ATPase localization or for proton pumping. The peripheral location of the protein, away from the interface of subunit a and the c-ring, leaves its function in the V-ATPase complex unclear. A BLAST search indicates that YPR170W-B is highly conserved in fungi. Reliable identification of possible animal or plant homologues of subunit f will require experiments with other organisms. The cryo-EM density for YPR170W-B (red–brown in Fig. 1) had insufficient resolution to build an atomic model; consequently this component of the complex was modelled as poly-alanine.

The asymmetric distribution of Glu residues around the c-ring, and how the ring interacts with subunit a, has implications for possible modes of proton translocation. The spacing between Glu residues in the c-ring has an increment of 36° between most Glu residues, but 18° between the Glu residues of subunits c'' and c₍₁₎ (Fig. 2b, purple and light pink), and 54° between the Glu residues of subunits c' and c'' (Fig. 2b, magenta and purple). As seen in this model, subunit c'' is positioned at the protonating cytoplasmic half-channel (Fig. 4a, blue arrow). In this orientation, subunit c₍₁₎ is in contact with subunit a at a possible mid-membrane origin for the luminal half-channel (Fig. 4a, red arrow). However, if the ring were rotated such that one of the c subunits (for example, subunit c₍₁₎) engages with the cytoplasmic half-channel (Fig. 4b, blue arrow) then the next c subunit (subunit c₍₂₎ in this example) would not be in contact with subunit a, making it unlikely that it is near a luminal half-channel (Fig. 4b, red arrow). Misalignment with any possible luminal half-channel is even more pronounced if subunit c' is engaged with the cytoplasmic half-channel (Fig. 4c, blue arrow), in which case the Glu from subunit c'' is extremely far from contact with subunit a (Fig. 4c, red arrow). This mismatch ensures that protonation and deprotonation cannot always occur simultaneously. Instead, the sequence of events for ATP-hydrolysis-driven proton pumping must be that rotation of the c-ring, caused by ATP hydrolysis in the V₁ region, breaks the contact between Arg735 and a deprotonated Glu residue,

dragging the Glu residue into the hydrophobic environment of the lipid bilayer (Fig. 4d). Thermodynamic analysis has shown that moving a deprotonated Glu residue from an aqueous environment into the middle of a hydrophobic lipid bilayer will force the residue to acquire a proton and become neutralized³⁰. This thermodynamic property of Glu residues ensures that the ring does not 'slip', allowing a deprotonated Glu to move into the lipid bilayer, which would uncouple ATP hydrolysis from proton pumping. Rotation of the c-ring brings a protonated Glu residue out of the lipid bilayer and into alignment with the putative luminal half-channel. As rotation of the ring continues, the Glu residue encounters Arg735, possibly in a different orientation than observed in this structure, causing deprotonation of the Glu residue via the luminal half-channel and resetting the c-ring for a subsequent protonation–deprotonation cycle. The conformation of Arg735 may also need to adapt to accommodate interaction with the Glu residue of subunit c'', compared to subunit c or c'. This sequence of events, in which only one Glu residue at a time interacts with a half-channel, explains how rotary ATPases can tolerate c-rings with variable²¹ and unequal distances between the c-subunits.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 June; accepted 6 September 2016.

Published online 24 October 2016.

- Zhao, J., Benlekbir, S. & Rubinstein, J. L. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* **521**, 241–245 (2015).
- Allegretti, M. et al. Horizontal membrane-intrinsic α -helices in the stator a-subunit of an F-type ATP synthase. *Nature* **521**, 237–240 (2015).
- Zhou, A. et al. Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. *eLife* **4**, e10180 (2015).
- Morales-Rios, E., Montgomery, M. G., Leslie, A. G. W. & Walker, J. E. Structure of ATP synthase from *Paracoccus denitrificans* determined by X-ray crystallography at 4.0 Å resolution. *Proc. Natl Acad. Sci. USA* **112**, 13231–13236 (2015).
- Schep, D. G., Zhao, J. & Rubinstein, J. L. Models for the a subunits of the *Thermus thermophilus* V/A-ATPase and *Saccharomyces cerevisiae* V-ATPase enzymes by cryo-EM and evolutionary covariance. *Proc. Natl Acad. Sci. USA* **113**, 3245–3250 (2016).
- Kane, P. M. Disassembly and reassembly of the yeast vacuolar H⁺-ATPase *in vivo*. *J. Biol. Chem.* **270**, 17025–17032 (1995).
- Sumner, J. P. et al. Regulation of plasma membrane V-ATPase activity by dissociation of peripheral subunits. *J. Biol. Chem.* **270**, 5649–5653 (1995).
- Couoh-Cardel, S., Milgrom, E. & Wilkens, S. Affinity purification and structural features of the yeast vacuolar ATPase Vo membrane sector. *J. Biol. Chem.* **290**, 27959–27971 (2015).
- Benlekbir, S., Bueler, S. A. & Rubinstein, J. L. Structure of the vacuolar-type ATPase from *Saccharomyces cerevisiae* at 11-Å resolution. *Nat. Struct. Mol. Biol.* **19**, 1356–1362 (2012).
- Qi, J. & Forgac, M. Function and subunit interactions of the N-terminal domain of subunit a (Vph1p) of the yeast V-ATPase. *J. Biol. Chem.* **283**, 19274–19282 (2008).
- Powell, B., Graham, L. A. & Stevens, T. H. Molecular characterization of the yeast vacuolar H⁺-ATPase proton pore. *J. Biol. Chem.* **275**, 23654–23660 (2000).
- Hirata, R., Graham, L. A., Takatsuki, A., Stevens, T. H. & Anraku, Y. *VMA11* and *VMA16* encode second and third proteolipid subunits of the *Saccharomyces cerevisiae* vacuolar membrane H⁺-ATPase. *J. Biol. Chem.* **272**, 4795–4803 (1997).
- Nishi, T., Kawasaki-Nishi, S. & Forgac, M. The first putative transmembrane segment of subunit c' (Vma16p) of the yeast V-ATPase is not necessary for function. *J. Biol. Chem.* **278**, 5821–5827 (2003).
- Couoh-Cardel, S., Hsueh, Y.-C., Wilkens, S. & Movileanu, L. Yeast V-ATPase proteolipid ring acts as a large-conductance transmembrane protein pore. *Sci. Rep.* **6**, 24774 (2016).
- Matthies, D. et al. High-resolution structure and mechanism of an F₁/V-hybrid rotor ring in a Na⁺-coupled ATP synthase. *Nat. Commun.* **5**, 5286 (2014).
- Vik, S. B. & Antonio, B. J. A mechanism of proton translocation by F₁F₀ ATP synthases suggested by double mutants of the a subunit. *J. Biol. Chem.* **269**, 30364–30369 (1994).
- Junge, W., Lill, H. & Engelbrecht, S. ATP synthase: an electrochemical transducer with rotatory mechanics. *Trends Biochem. Sci.* **22**, 420–423 (1997).
- Neale, C., Chakrabarti, N., Pomorski, P., Pai, E. F. & Pomès, R. Hydrophobic gating of ion permeation in magnesium channel CorA. *PLOS Comput. Biol.* **11**, e1004303 (2015).

19. Angevine, C. M., Herold, K. A. & Fillingame, R. H. Aqueous access pathways in subunit a of rotary ATP synthase extend to both sides of the membrane. *Proc. Natl Acad. Sci. USA* **100**, 13179–13183 (2003).
20. Toei, M., Toei, S. & Forgac, M. Definition of membrane topology and identification of residues important for transport in subunit a of the vacuolar ATPase. *J. Biol. Chem.* **286**, 35176–35186 (2011).
21. Kühlbrandt, W. & Davies, K. M. Rotary ATPases: a new twist to an ancient machine. *Trends Biochem. Sci.* **41**, 106–116 (2016).
22. Pellegrini-Calace, M., Maiwald, T. & Thornton, J. M. PoreWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLOS Comput. Biol.* **5**, e1000440 (2009).
23. Smart, O. S., Neduelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. P. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360, 376 (1996).
24. Kawasaki-Nishi, S., Nishi, T. & Forgac, M. Arg-735 of the 100-kDa subunit a of the yeast V-ATPase is essential for proton translocation. *Proc. Natl Acad. Sci. USA* **98**, 12397–12402 (2001).
25. Cain, B. D. & Simoni, R. D. Impaired proton conductivity resulting from mutations in the a subunit of F_1F_0 ATPase in *Escherichia coli*. *J. Biol. Chem.* **261**, 10043–10050 (1986).
26. DeCoursey, T. E. The voltage-gated proton channel: a riddle, wrapped in a mystery, inside an enigma. *Biochemistry* **54**, 3250–3268 (2015).
27. Bueler, S. A. & Rubinstein, J. L. Vma9p need not be associated with the yeast V-ATPase for fully-coupled proton pumping activity *in vitro*. *Biochemistry* **54**, 853–858 (2015).
28. Nelson, H. & Nelson, N. Disruption of genes encoding subunits of yeast vacuolar H^+ -ATPase causes conditional lethality. *Proc. Natl Acad. Sci. USA* **87**, 3503–3507 (1990).
29. Compton, M. A., Graham, L. A. & Stevens, T. H. Vma9p (subunit e) is an integral membrane V_0 subunit of the yeast V-ATPase. *J. Biol. Chem.* **281**, 15312–15319 (2006).
30. MacCallum, J. L., Bennett, W. F. D. & Tieleman, D. P. Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys. J.* **94**, 3393–3404 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank N. Grigorieff for providing access to the Titan Krios electron microscope and H. Urlaub for providing C.S. with access to mass spectrometry instrumentation while in Göttingen. We thank Z. Ripstein, P. Tieleman, R. Pomès, and J.-P. Julien for discussions and J. Zhao, J.-P. Julien and V. Kanelis for a critical reading of the manuscript. M.T.M.-J. was supported by a Postdoctoral Fellowship from the Canadian Institutes of Health Research (CIHR). C.V.R. is a Royal Society Professor and J.L.R. holds a Canada Research Chair. This work was supported by operating grant MOP81294 from the Canadian Institutes of Health Research (J.L.R.), Wellcome Trust grants WT008150 and WT099141 (C.V.R.), and European Research Council IMPRESS grant ERC268851 (C.V.R.).

Author Contributions M.T.M.-J. prepared yeast strains, purified protein, prepared cryo-EM specimens, performed 200 kV cryo-EM and Relion image analysis, and built the atomic models. A.R. collected and pre-processed data with the 300-kV microscope and performed image analysis with FREALIGN. C.S. performed the mass spectrometry analysis. S.A.B. prepared yeast strains and assisted with protein purification. S.B. assisted with cryo-EM specimen preparation and screening. C.V.R. supervised mass spectrometry experiments, and J.L.R. supervised the other aspects of the work and coordinated experiments. J.L.R. and M.T.M.-J. wrote the manuscript and prepared the figures with input from the other authors.

Author Information Cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-8363, EMD-8364, EMD-8367, and EMD-8409. The atomic model has been deposited in the Protein Data Bank under accession number 5TJ5. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L.R. (john.rubinstein@utoronto.ca).

Reviewer Information *Nature* thanks A. Hack, K. Parra, H. Saibil and J. Weber for their contribution to the peer review of this work.

METHODS

Yeast strains and protein purification. To isolate the *S. cerevisiae* V_0 complex, yeast strain CACY1, expressing Vph1p (subunit a) with a 3×FLAG-tag at its C terminus, was prepared by homologous recombination as described previously⁹. To confirm that YPR170W-B is a subunit of the V-ATPase, an *S. cerevisiae* strain was prepared with a 3×FLAG-tag fused to the C terminus of the uncharacterized open reading frame YPR170W-B (*Saccharomyces* genome database identifier S000028515) in the protease-deficient background strain MM93, producing the strain SABY87. To purify intact V-ATPase, we used strain SABY31, which bears a 3×FLAG-tag on Vma1p and with the *STV1* gene deleted⁹. YPR170W-B was deleted in strain CACY1 by homologous recombination with the NatR cassette to produce strain MMJY1 as previously described⁹. Western blotting against 3×FLAG-tagged proteins was done with the monoclonal anti-FLAG antibody M2 (Sigma). For protein purification, yeast in YPD medium were grown in an 11-l BioFlo fermenter (New Brunswick Scientific) for 60 h in order to induce nutrient starvation and dissociation of the V_1 and V_0 regions of the complex. V_0 complex solubilized with dodecylmaltoside (DDM) was purified via the 3×FLAG-tag with M2 affinity matrix (Sigma), following the same protocol used for the intact V-ATPase^{3,9}. After purification in elution buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.02% (w/v) DDM, 50 µg/ml 3×FLAG peptide), the V_0 complex was mixed with amphipol A8-35 (Anatrace) at a protein:amphipol ratio of 1:10 (w/w) with gentle agitation for 1 h. Detergent was removed with 15 mg/ml Bio-Beads SM-2 (Bio-rad) at 4 °C overnight. The sample was purified further with a Superdex 200 column previously equilibrated with buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl). Protein from the chromatographic peak corresponding to the V_0 complex was collected and concentrated to 2.5 mg/ml with a 100-kDa MWCO Amicon concentrator (Millipore) and bafilomycin A1 was added to 10 µM (Santa Cruz Biotechnology) before further analysis.

LC-MS/MS and database search. Subunits of the V_0 preparation were separated by SDS-PAGE and regions of the gel were excised at positions where small transmembrane α -helical hairpin subunits are expected. Proteins were digested in the gel as described previously³¹ with trypsin and chymotrypsin at 37 °C and 25 °C, respectively. For combined tryptic and chymotryptic digestion, proteins were digested with trypsin at 37 °C for 3 h before chymotrypsin was added and the sample was incubated at 25 °C overnight. Peptides dissolved in 2% (v/v) acetonitrile and 0.1% (v/v) formic acid were separated by nano-flow liquid chromatography (Dionex UltiMate 3000 RSLC, Thermo scientific; mobile phase A: 0.1% (v/v) formic acid; mobile phase B: 80% (v/v) acetonitrile, 0.08% (v/v) formic acid). Peptides were then loaded onto a trap column (Reprosil C18, 100 µm inner diameter, particle size 5 µm; Dr. Maisch GmbH, prepared in-house) and separated with a flow rate of 300 nl/min on an analytical C18 capillary column (Reprosil C18, 75 µm inner diameter, particle size 1.9 µm, 27–28 cm; Dr. Maisch GmbH, prepared in-house), with a gradient of 5–90% (v/v) mobile phase B over 46 min. Separated peptides were directly eluted into a Orbitrap Fusion Tribrid Mass Spectrometer (Thermo scientific). Typical mass spectrometric conditions were: spray voltage of 2.3 kV; capillary temperature of 275 °C; collision energy of 30%; activation Q of 0.25. The Orbitrap Fusion Tribrid Mass Spectrometer was operated in data-dependent mode. Survey full-scan MS spectra were acquired in the orbitrap (m/z 380–1,500) with a resolution of 120,000 and an automatic gain control (AGC) target at 400,000. The top-10 most intense ions were selected for higher-energy collisional dissociation MS/MS fragmentation in the orbitrap at a resolution of 30,000 and an AGC target of 1,200 and with a first m/z of 110. Dynamic exclusion of previously selected ions was set to 30 s. Only ions with charge states 2–7 were selected. Singly and doubly charged ions, as well as ions with an unrecognized charge state, were also excluded. Internal calibration of the orbitrap was performed with the lock mass option (lock mass, m/z 445.120025)³². Raw files were converted into mgf files using Proteome Discoverer (Thermo scientific). Mgf files were searched against Uniprot_Yeast database (23,481 sequences) using Mascot search engine v2.03.2002 (Matrix Science). Search parameters were: peptide mass tolerance, 10 p.p.m.; fragment mass tolerance, 0.6 Da; enzyme, trypsin; variable modifications, carbamidomethylation (cysteine) and oxidation (methionine).

Cryo-EM and image analysis. V_0 complex (2.5 µl) in amphipol was applied to nano-fabricated, holey-gold-coated EM grids³³ previously glow-discharged in air for 15 s, blotted for 15 s, and then plunge-frozen with a modified Vitrobot grid-preparation device (FEI Company) in a mixture of liquid ethane and propane at liquid-nitrogen temperature³⁴. Cryo-EM grid preparation conditions were optimized and a small dataset consisting of 1,082 images obtained with a field-emission Tecnai F20 electron microscope (FEI Company) operating at 200 kV, with images recorded on a Gatan K2 Summit (Gatan Inc.) direct detector device camera (counting mode, 2 frames/s, 15 s, 1.45 Å/pixel, 1.2 e[−]/Å²/frame). An initial model for the V_0 complex was generated by manually segmenting a map of the intact V-ATPase¹ with UCSF Chimera³⁵ and low-pass filtering to 30 Å. Image analysis with Alignframes_{lmbf}³⁶, CTFIND3 (ref. 37). Alignparts_{lmbf}³⁶, magnification anisotropy

correction³⁸ and Relion 1.3 (ref. 39) produced a preliminary map at 6.8-Å resolution from 39,384 particle images. The structure of the V_0 complex in DDM was also determined from yeast strains CACY1 (wild type) and MMJY1 (Δ YPR170W-B). These specimens were prepared with nanofabricated holey-carbon-coated EM grids subjected to glow-discharge for 2 min and blotted for 20 s. 3D maps were calculated at 8.3-Å resolution from 43,184 particle images (for CACY1) and at 8.7-Å resolution from 44,468 particle images (for MMJY1).

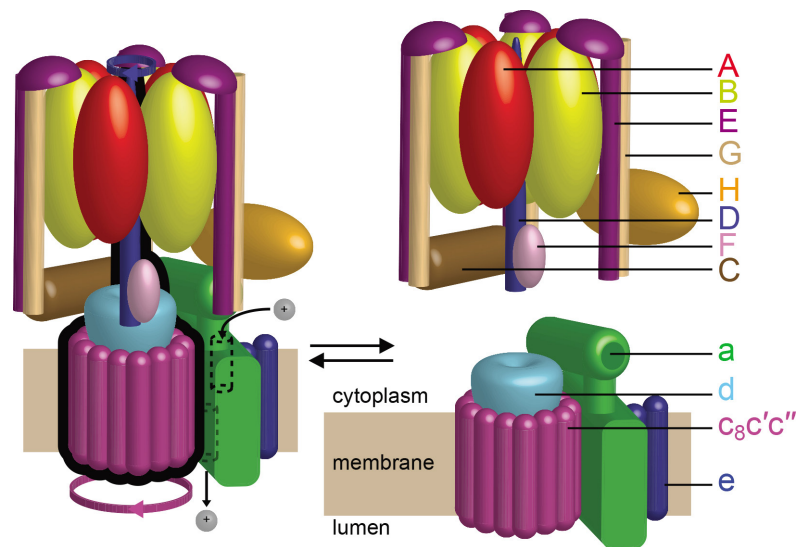
For high-resolution image acquisition, grids were sent in a cryogenic specimen shipper to the HHMI Janelia Research Campus where they were imaged with a Titan Krios electron microscope (FEI Company). Micrographs were recorded from a single grid with the microscope operated at 300 kV using parallel illumination at 4.8 e[−]/Å²/s of a 1.58-µm diameter region of the grid from a 70-µm objective aperture. Images were recorded with a K2 Summit direct detector device camera operating in super-resolution mode with a nominal magnification of 37,000×. With no specimen present in the optical path, the rate of exposure of the detector was 3 e[−]/pixel/s. Dose-fractionated exposures of 21 s were recorded as movies with 70 frames, so that selected specimen areas were exposed to a total of 100 e[−]/Å². Data collection was automated with SerialEM⁴⁰. A previously measured magnification anisotropy was corrected with the program mag_distortion_correct⁴¹, leading to a super-resolution pixel size of 0.3885 Å. Frames were down-sampled to a pixel size of 1.554 Å by Fourier-space cropping and aligned with each other using Unblur⁴². Probably because of the high magnification and small field of view, no apparent advantage was detected for correcting images for individual particle motion³⁶. Defocus parameters were estimated with CTFIND4 (ref. 43) from the average of amplitude spectra of sums of 3 unaligned frames resampled to a pixel size of 1.94 Å. 657,975 particle images were automatically selected with Relion from 4,365 aligned and averaged movies, and extracted into 200 × 200 pixel boxes. 2D classification with Relion reduced the dataset to 462,842 particle images.

The map showing the V_0 complex lacking subunit d was obtained by 3D classification and refinement with Relion (ref. 39). The 6.8-Å resolution map of the V_0 complex in amphipol from the Tecnai F20 dataset was used as a starting reference for aligning the full dataset of particle images with FREALIGN⁴⁴. In all subsequent processing steps, information beyond 6-Å resolution was excluded to prevent over-fitting of noise at higher resolutions. Initial particle orientation parameters were obtained by 1 cycle of grid search in FREALIGN mode 3, using an angular step size of 3.4°. Following this cycle, and before every subsequent cycle, the solvent and amphipol regions of the map were low-pass-filtered to 30 Å to reduce the fitting of noise from solvent and amphipol densities⁴⁵. Final particle-orientation parameters were obtained after 19 cycles of local refinement (mode 1), one cycle of mode 3 exhaustive search, and 11 further cycles of local refinement. The final 3D map was calculated from particle images padded with zeroes from 200 × 200 to 400 × 400 pixels to mitigate CTF aliasing effects⁴⁶ and was then cropped to a 200 × 200 × 200 voxel volume. Overall resolution of the map was estimated by Fourier shell correlation to be 3.9 Å. Local resolution variability was estimated with blocres⁴⁷.

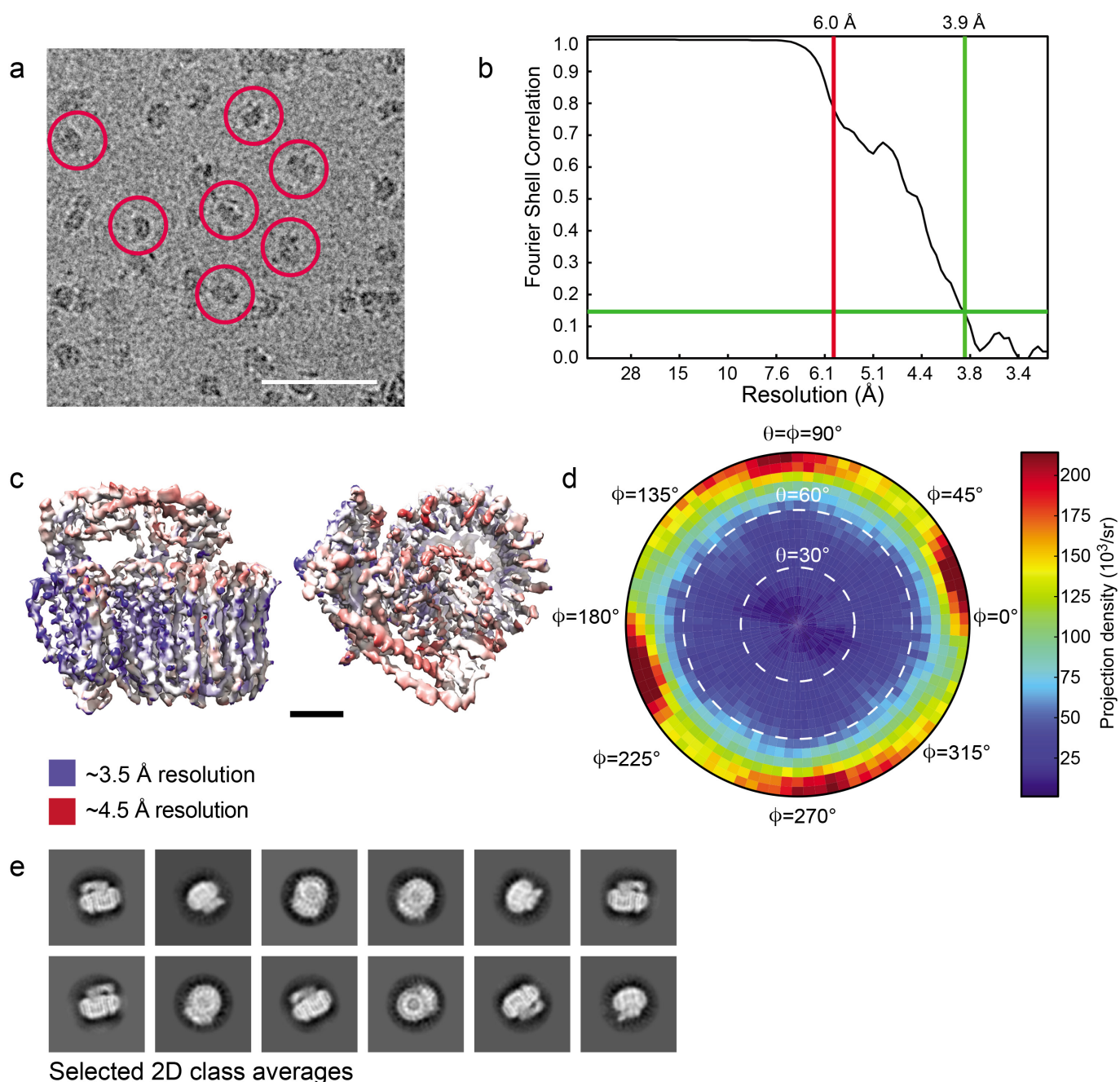
Model building. Most regions of the map, particularly membrane-embedded α -helices, had sufficient details to allow *de novo* model building. Initial models for subunits c, c', c'', d, and the N-terminal domain of subunit a were generated with the SWISS-MODEL server⁴⁸ using a model of intact V-ATPase from cryo-EM (Protein Databank accession number 3J9V) as a template. The initial model for the membrane-embedded domain of subunit a was from the cryo-EM and evolutionary covariance analysis of that protein (Protein Databank accession number 5IIM)⁵. Initial models for subunits e and f were generated manually in Coot⁴⁹. Initial models were fit into density as rigid bodies with UCSF Chimera³⁵. Final models were built with successive rounds of real-space refinement in Phenix⁵⁰ and manual model building in Coot and gave an EMRinger score of 2.0 for the entire model⁵¹, which is superior to the typical score of 1.0 for a 4-Å map. 93.2, 6.3, and 0.5% of residues were in preferred, allowed, and disallowed regions of the Ramachandran plot, respectively, with no Ramachandran outliers in the α -helical regions of the model. Where no corresponding densities were observed, side chains were deleted while maintaining the residue identity. Subunit f and the N-terminal domain of subunit a were modelled entirely as poly-alanine.

- Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–858 (1996).
- Olsen, J. V. *et al.* Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
- Marr, C. R., Benlekhir, S. & Rubinstein, J. L. Fabrication of carbon films with ~500 nm holes for cryo-EM with a direct detector device. *J. Struct. Biol.* **185**, 42–47 (2014).
- Tivol, W. F., Briegel, A. & Jensen, G. J. An improved cryogen for plunge freezing. *Microsc. Microanal.* **14**, 375–379 (2008).
- Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).

36. Rubinstein, J. L. & Brubaker, M. A. Alignment of cryo-EM movies of individual particles by optimization of image translations. *J. Struct. Biol.* **192**, 188–195 (2015).
37. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
38. Zhao, J., Brubaker, M. A., Benlekhir, S. & Rubinstein, J. L. Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. *J. Struct. Biol.* **192**, 209–215 (2015).
39. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
40. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
41. Grant, T. & Grigorieff, N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* **192**, 204–208 (2015).
42. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
43. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
44. Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
45. Grigorieff, N. FREALIGN: an exploratory tool for single-particle Cryo-EM. *Methods Enzymol.* **579**, 191–226 (2016).
46. Penczek, P. A. *et al.* CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* **140**, 9–19 (2014).
47. Cardone, G., Heymann, J. B. & Steven, A. C. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **184**, 226–236 (2013).
48. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
50. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
51. Barad, B. A. *et al.* EMRinger: side chain-directed model and map validation for 3D cryo-electron microscopy. *Nat. Methods* **12**, 943–946 (2015).

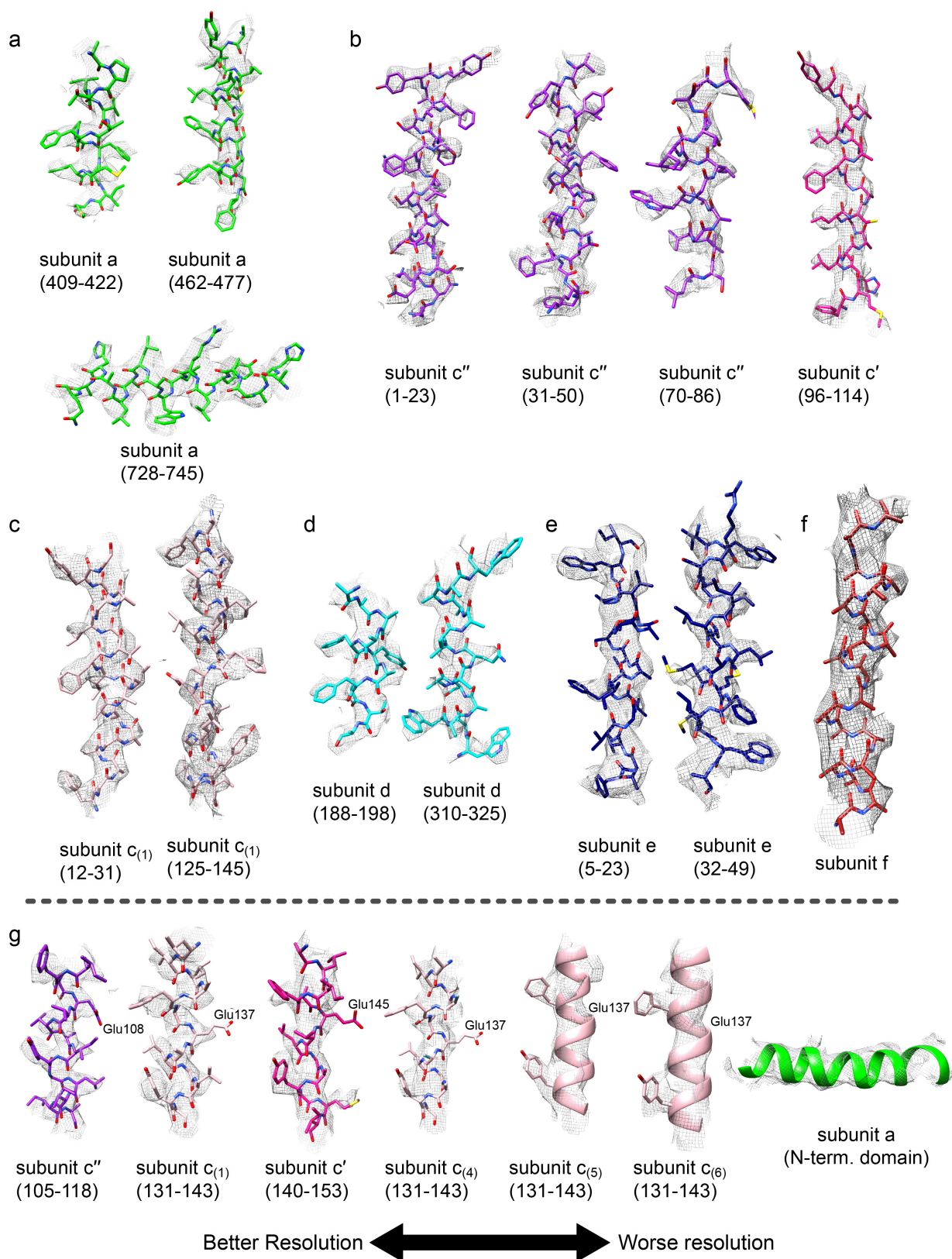


Extended Data Figure 1 | Subunit composition of the intact V-ATPase and dissociated V₁ and V₀ regions. The rotor is outlined in black and the two half-channels in the V₀ region are indicated with dashed lines. The intact V-ATPase (left) dissociates into the auto-inhibited V₁ and V₀ complexes upon nutrient starvation. Figure adapted from ref. 1.

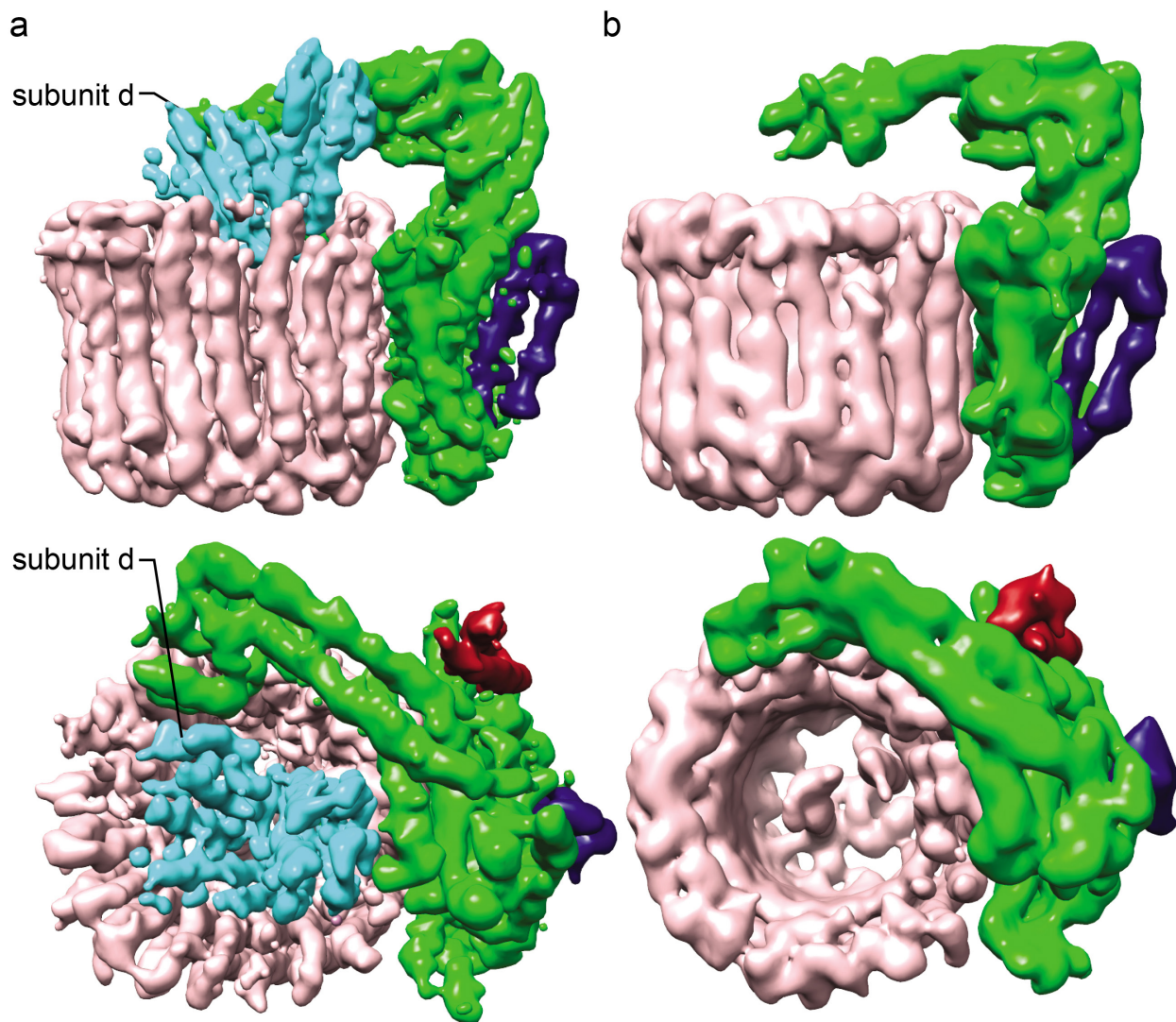


Extended Data Figure 2 | Cryo-EM map generation. **a**, An example micrograph with protein particles circled in red. Scale bar, 500 Å. **b**, Fourier shell correlation (FSC) curve. The highest-resolution information used in image alignment (6 Å) and the overall resolution

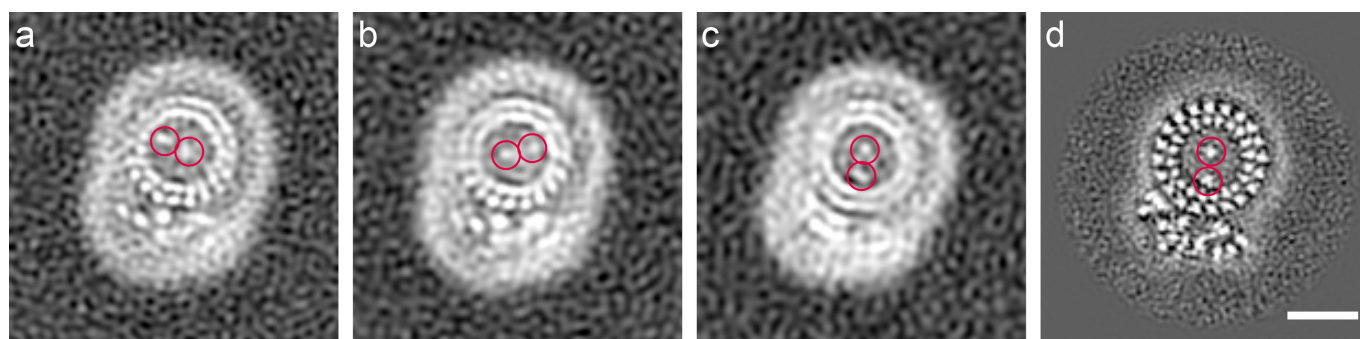
of the map at FSC = 0.143 (3.9 Å) are indicated. **c**, Local resolution assessment. Scale bar, 25 Å. **d**, Image orientation distribution. **e**, Example 2D class average images.



Extended Data Figure 3 | Model building. **a–f**, Example regions of the atomic model built for subunits a (**a**), c'' and c' (**b**), c₍₁₎ (**c**), d (**d**), e (**e**), and f (**f**). **g**, The different α -helices from the c-ring bearing conserved Glu residues show variable resolution. An α -helix from the N-terminal domain of subunit a has poor resolution. Residue numbers are shown in brackets.



Extended Data Figure 4 | V_O complex lacking subunit d. **a**, The V_O complex map from all of the particle images shows subunit d. **b**, V_O complex map from a 3D class, containing 24,744 particle images, that lacks subunit d was determined at 7.8-Å resolution. Scale bar, 25 Å.



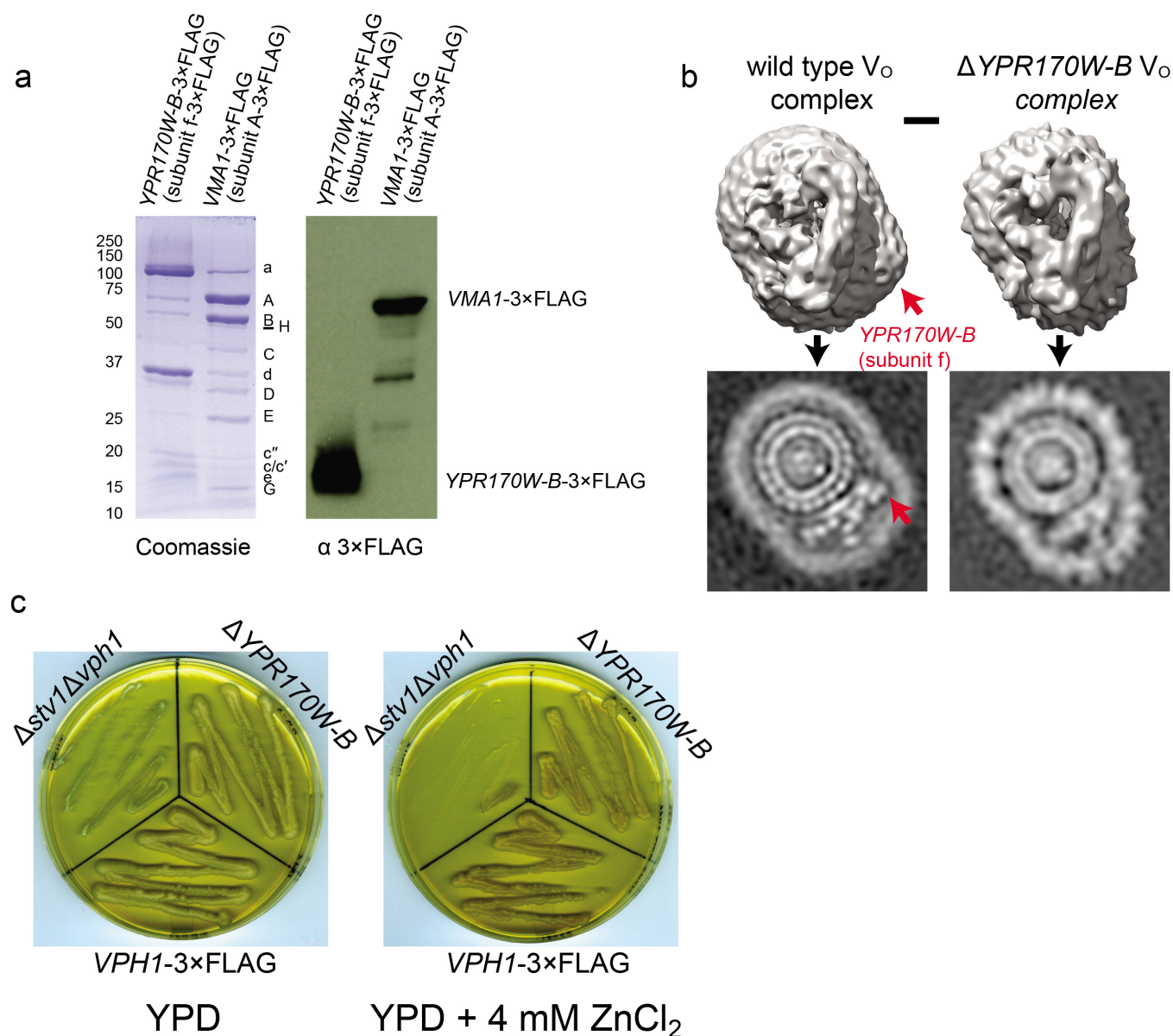
Intact V-ATPase
Rotational State 1
EMD-6284

Intact V-ATPase
Rotational State 2
EMD-6285

Intact V-ATPase
Rotational State 3
EMD-6286

V_O complex map

Extended Data Figure 5 | V_O complex is in rotational state 3. **a–c**, rotational states 1, 2, and 3 of the intact V-ATPase show the two α helices of subunit c'' within the c -ring¹. **d**, The two α -helices of subunit c'' within the c -ring show the ring to be in the same orientation as in rotational state 3 of the intact V-ATPase. Scale bar, 25 Å.



Extended Data Figure 6 | Identification of subunit f. **a**, SDS-PAGE gel (left) and western blot (right) against a 3×FLAG-tag for the affinity purification of 3×FLAG-tagged YPR170W-B (subunit f) and Vma1p (subunit A) show that both proteins are components of the V-ATPase. **b**, Surface-rendered 3D maps (upper) and map cross-sections (lower) showing the wild-type V_O complex (left) and the V_O complex from a

yeast strain with the YPR170W-B gene deleted (right). Density from YPR170W-B is indicated with a red arrow. Scale bar, 25 Å. **c**, Yeast strains with the *STV1* and *VPH1* genes deleted, the *STV1* and YPR170W-B gene deleted, and only *STV1* gene deleted were grown on both YPD medium (left) and YPD medium with zinc (right), demonstrating that deletion of YPR170W-B does not cause the VMA phenotype.

CORRECTIONS & AMENDMENTS

CORRIGENDUM

doi:10.1038/nature19335

Corrigendum: Lytic to temperate switching of viral communities

B. Knowles, C. B. Silveira, B. A. Bailey, K. Barott, V. A. Cantu, A. G. Cobián-Güemes, F. H. Coutinho, E. A. Dinsdale, B. Felts, K. A. Furby, E. E. George, K. T. Green, G. B. Gregoracci, A. F. Haas, J. M. Haggerty, E. R. Hester, N. Hisakawa, L. W. Kelly, Y. W. Lim, M. Little, A. Luque, T. McDole-Somera, K. McNair, L. S. de Oliveira, S. D. Quistad, N. L. Robinett, E. Sala, P. Salamon, S. E. Sanchez, S. Sandin, G. G. Z. Silva, J. Smith, C. Sullivan, C. Thompson, M. J. A. Vermeij, M. Youle, C. Young, B. Zgliczynski, R. Brainard, R. A. Edwards, J. Nulton, F. Thompson & F. Rohwer

Nature **531**, 466–470 (2016); doi:10.1038/nature17193

In this Article, the ‘Predator–prey modelling’ section of the Methods shows Lotka–Volterra equations. Although these equations are meant to present a basic Lotka–Volterra model, the term ‘ N/K ’ in the second equation was inadvertently introduced during proofing, which makes the equations reflect the Piggyback-the-Winner model rather than basic Lotka–Volterra. The equation should have read:

$$\delta V / \delta t = (\beta \cdot \phi \cdot N \cdot V) - (m \cdot V)$$

rather than:

$$\delta V / \delta t = (\beta \cdot \phi \cdot N / K \cdot N \cdot V) - (m \cdot V)$$

This change does not affect the conclusions of the paper, and has been corrected online.

CORRIGENDUM

doi:10.1038/nature19837

Corrigendum: Noncanonical autophagy inhibits the autoinflammatory, lupus-like response to dying cells

Jennifer Martinez, Larissa D. Cunha, Sunmin Park, Mao Yang, Qun Lu, Robert Orchard, Quan-Zhen Li, Mei Yan, Laura Janke, Cliff Guy, Andreas Linkermann, Herbert W. Virgin & Douglas R. Green

Nature 533, 115–119 (2016); doi:10.1038/nature17950

In Fig. 2a of this Letter, during the preparation of the final figures, the Cre⁻ Atg5^{fl/fl} representative image was inadvertently duplicated in lieu of the Nox2^{+/+} representative image. In Extended Data Fig. 2d, the Cre⁺ ATG7^{fl/fl} representative image for Ki67 immunohistochemical staining was inadvertently duplicated in lieu of the NOX2^{-/-} representative image. We sincerely apologize for these errors. The corrected image for Fig. 2a is shown in Fig. 1 to this Corrigendum, and the corrected Extended Data Fig. 9 is shown in the Supplementary Information. Quantifications in Fig. 2b are unaffected by this error.

Supplementary Information is available in the online version of the Corrigendum.

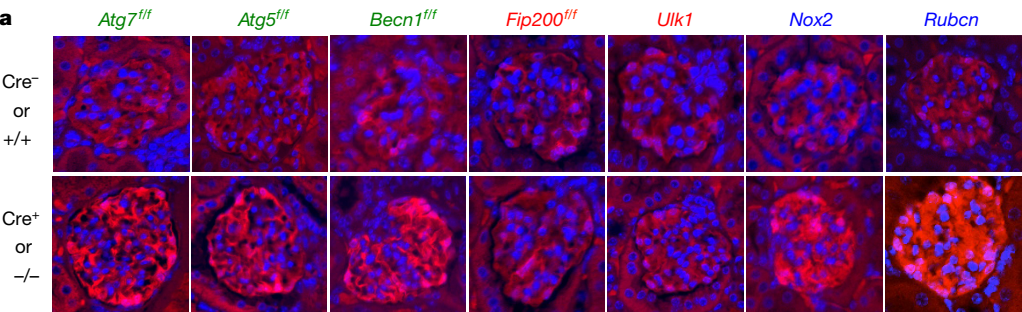


Figure 1 | This shows the corrected Fig. 2a from the original Letter.

MINING THE SECRETS OF COLLEGE SYLLABUSES

The creators of the Open Syllabus Project hope that sharing data can both improve and reward teaching.

ILLUSTRATION BY THE PROJECT TWINS



BY ANNA NOWOGRODZKI

Despite a growing movement to glean insights from scholarly materials that are available online — from articles and data sets to conference presentations and lectures — one kind of academic document remains little examined. And that is the syllabus: a document that lays out the reading materials, topics and expectations of college courses.

That, at least, was the case until January this year, when data scientists, sociologists and digital-humanities researchers at Columbia University in New York City launched a tool called the Open Syllabus Explorer. This integrates more than 1 million publicly available syllabuses and lays open their data in a conveniently searchable format. A version containing three times as many syllabuses is scheduled to launch in January 2017.

The team behind the tool, the Open Syllabus Project (OSP), hope to nudge universities towards making more syllabuses public. They argue that doing so could aid textbook authors, instructors and course developers, and would reward the design of effective teaching materials, which is largely overlooked by conventional measures of academic effort.

“Syllabuses are among the most important documents written by scholars which are not yet widely shared, and they ought to be,” says Peter Suber, director of the Harvard Open Access Project and the Harvard Office for Scholarly Communication in Cambridge, Massachusetts, who serves on the OSP advisory board. “They reflect serious scholarly judgements about what’s worth teaching.”

Such judgements can be welcome news to textbook authors. Stuart Russell, a computer scientist at the University of California,

Berkeley, didn’t realize until *Nature* interviewed him for this article that his 1995 book *Artificial Intelligence* (Prentice Hall), co-authored with Peter Norvig, was the most highly assigned text in the field of computer science. “I was definitely surprised,” he says.

Beyond stoking professional pride, such information could strengthen tenure and promotion packages. Authoring a textbook, no matter how useful and informative it might be, generally yields few citations in scholarly papers, so its academic impact is likely to be low. The OSP could help to shift the balance. “We’re at a point in time when I think faculty have to take more ownership of their whole record of scholarship, of impact, of influence,” says Amy Brand, director of the MIT Press. Hard data on syllabus usage, she says, could empower faculty members “to tell their own story about what their work is doing in the world”. ▶

► At present, Open Syllabus Explorer searches more than 1 million syllabuses dating back to 2000, cross-referenced with 20 million texts, to produce data on how often a text is taught. Users can search those data by author, title, institution and academic discipline. The tool also reports which textbooks are commonly used together, and ranks each text on how frequently it is taught (see ‘Top texts’).

An updated version, due to become available on 21 January 2017, the Explorer’s first anniversary, will feature 3 million syllabuses cross-referenced with about 150 million texts; these will include titles from the arXiv preprint server, CrossRef and the Virtual International Authority File — which links together identical bibliographic records from different national-library catalogues. The update will include new search options, such as the ability to search by date or type of institution, says Joe Karaganis, the OSP’s project director. The new version will also incorporate better Canadian and UK data, information about where to find materials and, eventually, full-text syllabuses, if the authors have given permission to reproduce them.

“We have some big ambitions,” Karaganis says. “All the techniques are very crude at present but they’re all improvable, and the data science is only getting better.”

FISHING FOR CITATIONS

The OSP is based at the American Assembly, a public-policy institute at Columbia University, and is funded by the Sloan Foundation and the Arcadia Fund. It was inspired by a search engine called Syllabus Finder, which scraped the public web for syllabuses from 2002 (the year it was built) until 2009. That tool was created by Dan Cohen, then a historian at George Mason University in Fairfax, Virginia, who is now executive director of the Digital Public Library of America. It amassed what Cohen says was then the largest collection of syllabuses ever assembled, comprising about 1 million documents. He released the URLs as a database in 2011.

Unlike the OSP, Cohen’s tool provided links to the full text of each syllabus. But it included only courses run up to 2009, when he had to retire the tool because of changes to Google’s programming interface — a move that vexed Cohen’s colleagues, including his wife, an early-childhood educator. “I still get e-mails begging me to turn the Syllabus Finder back on,” he says.

When the OSP began in 2014, the team built tools to scrape the public Internet — including the links used by Cohen, who had lost a portion of the data owing to a coding error. But, as Cohen was, the team is limited to publicly accessible syllabuses: about 6 million of an estimated 80 million to 120 million syllabuses in the United States alone, by Karaganis’s reckoning. Syllabuses sealed behind the walls of private course-management software, such as Blackboard, remain out of reach. “Columbia, for example, is sitting on 80,000 syllabuses from the last

TOP TEXTS

The 5 most-taught scientific texts according to OSP.

Textbook (author)	Syllabuses
<i>Biology: Concepts and Connections</i> (N. A. Campbell et al.)	2,196
<i>Fundamentals of Anatomy and Physiology</i> (F. Martini et al.)	752
<i>Chemistry</i> (R. Chang)	612
<i>Human Anatomy & Physiology</i> (E. N. Marieb and K. Hoehn)	605
<i>Human Anatomy</i> (E. N. Marieb et al.)	591

Data filtered by fields: Astronomy and Astrophysics, Biology, Chemistry, Computer Science, Earth Sciences, Engineering, Psychology, Sociology.

12 or 13 years,” says Karaganis. “A large state school could have two, three times that.”

The OSP team then had to build tools to extract what those syllabuses contained. Citations, for instance, had no consistent structure, says David McClure, the project’s technical director. The tool searched for titles by cross-referencing each syllabus against a database of 20 million titles — 11 million from Harvard LibraryCloud and 9 million from JSTOR. A matching title and author counted as a citation. “We built in different techniques for allowing fuzziness, like allowing the word ‘by’ in between the author and title,” says McClure.

A NEW METRIC

The OSP distils those data down to a single metric called the teaching score, which indicates how often a text is assigned in syllabuses. It can take any value from 1 (rarely taught) to 100 (frequently taught).

According to Suber, teaching scores are an alternative to conventional metrics of scholarly impact. They reflect the burgeoning ‘alternative metrics’ ethos, which aims to quantify the whole of a person’s research output. “I think this teaching score can take part in the new alt-metrics movement and give us a more sensitive measurement of the impact of texts,” he says.

Already, a handful of researchers and universities are using the data to do just that. The University of Kentucky in Lexington issued a press release when it discovered that a paper by Edward Morris, one of its faculty members, ranked 46 out of 13,225 sociology-related texts. It now ranks 371 out of 53,177, and Morris plans to use the figure to support his promotion to full professor.

US universities aren’t the only ones paying attention. Most of the roughly 1,000 visits to the OSP each day are from the United States, says Karaganis, but significant traffic comes from Ukraine, Russia and Egypt as well.

Other researchers have used the data to compile lists of widely taught graphic novels and comics, for instance, or to quantify the fraction of frequently taught sociology texts authored by women. Melanie Martin, a post-doc at Yale University in New Haven, Connecticut, used the Syllabus Explorer to identify

the most commonly taught texts in her field, evolutionary anthropology. But, because there is no way to search the database by subfield — for instance, limiting biology results to such subdisciplines as neuroscience or genomics — she had to scan the 16,000 anthropology titles manually. “Without better filtering, I think it’s limited,” she says.

BUILDING ON PEER EXPERTISE

Another possible application of OSP data involves course design. By enabling faculty members — particularly junior ones — to build on the knowledge of their peers, the OSP could help them to teach more creatively, such as by identifying new ways to present teaching material. “This could go a long way to improving the quality of instruction,” says Russell. It would also improve efficiency, leaving faculty members more time for other activities such as research and mentoring.

However, it is important not to over-interpret the data, says Lisa Janicke Hinchliffe, a specialist in information literacy at the University of Illinois at Urbana–Champaign. The project’s sample set might not be a good proxy for all syllabuses, even at a particular institution. For instance, the second most-assigned text at Harvard, according to the Explorer, is ‘Letter from Birmingham Jail’ by Martin Luther King Jr. But about 80% of the OSP’s Harvard syllabuses come from the John F. Kennedy School of Government, Karaganis says (although the OSP doesn’t publicly list its sources in this much detail). So it’s not possible to conclude how popular this text is at Harvard overall.

For Hinchliffe, the value of the OSP lies in its ability to reveal the breadth of resources that instructors use. “I don’t need a definitive ‘These are the top-six taught books,’” she says. “I want to see the variety.”

Such information could go a long way towards simplifying course design, a notoriously time-consuming process. Just ask Suber, who has been teaching philosophy for 21 years. “Whenever I knew a new course was coming, I would try to start preparing it at least a year in advance,” he says. “Writing 40 lectures is a huge job; it’s harder than writing a book.”

The OSP’s data could ease that burden. Plus, says Suber, the data are fun to explore, sometimes revealing unexpected pairings. His legal philosophy text, *The Case of the Speluncean Explorers* (Routledge, 1998), for instance, has been taught alongside Sappho’s lyric poetry. “There are partners or juxtapositions that I never would have guessed,” he says. ■

CORRECTION

The Toolbox article ‘Democratic databases: science on GitHub’ (*Nature* **538**, 127–128; 2016) misstated how the Git software records changes in files. It does in fact maintain multiple versions of the files.

CAREERS

GENDER Fear of bias pushes women from STEM fields **p.129**

DRUGMAKERS Research jobs cut in the United Kingdom **p.129**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

PARTICK GEORGE/GETTY



POSTGRADUATE STUDIES

Find the best fit

Choices for doctoral programmes can seem endless, so look for one that matches your interests and personality.

BY KENDALL POWELL

There's much to consider when you're trying to choose the university and programme for your science PhD. But the main reason for your selection must be that it suits you — not that you don't know what else to do, not the institution's or department's reputation, not that a star researcher in your field is a faculty member there. Getting a PhD is hard enough, says Bruce Horazdovsky, associate dean for the Mayo Graduate School in Rochester, Minnesota. You don't want to make it harder by being

"miserable while you're doing it", he says. "You have to be engaged and like what you are doing. The best programme in the country is the one that best fits you."

How do you find that best fit? Prospective doctoral students will need to consider several factors and compare programmes and schools. Deciding which universities to apply to means identifying programmes that match your research interests and personality. You will need to evaluate how the school approaches career and professional development for its graduate students, and how its alumni fare after achieving their PhDs.

Ultimately, the school you select will be the launch pad for your scientific career.

Before you look at schools, you should have a clear idea of your chosen subfield of study. "Even at this stage, students ought to be thinking about what sort of specialization they want to do," says David Bogle, pro-vice-provost of the doctoral school at University College London. He notes that a physics programme, for instance, could be great for astrophysics and string-theory research but offer nothing on materials science.

Although it's not necessary to narrow down fields too specifically, it is imperative to find a programme that has at least several faculty members who are doing research that excites you, says Bogle, who chairs the League of European Research Universities' doctoral studies community in Leuven, Belgium. He advises students to look, not for a single high-profile researcher, but rather for a strong research environment with several professors working in similar areas.

To get started, applicants can generally find descriptions of a school's research programmes and faculty members on the institution's website. Sometimes, more information is available: the European School of Molecular Medicine (SEMM), a graduate programme shared between two universities and three research centres in Milan and Naples, Italy, publishes an annual list of faculty members who are taking new students in the coming year. Other institutions may publish similar material.

Group websites can also give applicants a feel for the size and culture of a laboratory. Look for photos of lab outings or celebrations, for announcements of student achievements and publications, and for other evidence that graduate students drive much of the research in the group. Applicants should also look up a lab group's latest research publications to get an idea of its members' current interests and to see how well and how often students in the lab are publishing papers. "If the publications coming out of a lab are numerous and high quality, you can be pretty sure that you will get published by the end of your PhD" — which is essential for success after graduation, says Francesca Fiore, coordinator of the SEMM graduate office in Milan.

Applicants should also seek advice and guidance from their undergraduate or master's advisers to generate a shortlist of potential programmes. "Come talk to me," says Andreas Berlind, an astrophysicist at Vanderbilt ►

► University in Nashville, Tennessee. “Let me help you make that initial list — it will save you a lot of time.” Advisers, he says, have enough deep knowledge of their field and its subfields to know which programmes tie in with the subjects a student is passionate about; they should also know where other researchers in that subfield are doing good work. For extra connections with the research world, applicants should try to attend large scientific conferences in their subfield; these often have travel fellowships so that undergraduates can attend.

Students should also check online resources such as the US National Research Mentoring Network (nrmnet.net) and Facebook sites such as Equity Einstein, a group dedicated to making physics and astronomy more inclusive. Such resources will help students to connect with established researchers who can offer advice on training. They should also contact current graduate students in their subfield to learn about programmes’ reputations. Applicants should not be shy about doing this, says Berlind, who is also Vanderbilt’s director of graduate studies in astrophysics. It is the best way to get honest answers about the culture and atmosphere in a programme, he adds (see ‘The value of hindsight’). Fiore also encourages correspondence with current students, especially for applicants who are pondering studying abroad. “Find someone from your home country,” if possible, she says, so that you can discuss their experience in your native language.

Prospective students should never pin their hopes on working with one particular professor, because that person may not be taking students, may move away or might

be a terrible fit as a mentor. If several faculty members are working in a similar area, the student has a better chance of landing a spot in one of those labs. Identify and contact at least two researchers, and ideally more, whom you’d like to do a PhD with, counsels Pamela McLean, director of neurobiology at Mayo Graduate School in Jacksonville, Florida. When you e-mail them, you can let them know of your interest in their work and find out whether they are taking on doctoral students in the next year. “A lot of times it will also strengthen your application,” she says. “Those names are often forwarded to admissions directors, and someone who has taken the initiative gets bonus points.”

SHOW ME THE DATA

More programmes are publishing data on their websites about their graduate students, including the average time taken to achieve a PhD. Students should pay particular attention to this: anything much more than five years for US programmes or three for UK programmes can indicate that students are languishing in labs as labourers. Some institutions provide data on their graduates’ career choices — the University of California, San Francisco, posts outcome data for most of its graduate-division programmes (go.nature.com/2dnyy89). It’s unusual for these data to be long-term enough to give a realistic picture of what all PhD holders are doing ten years after earning their degree, but it is still useful to scan such listings to see if doctoral graduates are ending up in careers that applicants consider desirable. “If they’re not there, that’s a bad sign that the department doesn’t see it as a priority to advertise

how well students are doing,” says Berlind.

Applicants should also determine whether they want to work on fundamental questions or do applied research. Students interested in the latter should seek programmes with strong ties to high-tech companies, the aerospace industry or hospitals, if their passion lies in those areas. For example, the Mayo Graduate School is spread across three large medical campuses in Minnesota, Florida and Arizona.

Students should also give some thought to the overall structure and organization of graduate programmes; these can be small and based in single departments or wide-ranging and interdisciplinary. Umbrella programmes (sometimes called structured programmes in the United Kingdom) pull in faculty members across several departments or campuses. These are in contrast to more conventional, single-department programmes, and in many cases they offer numerous labs and more options for cross-disciplinary studies. But what they make up for in quantity, they may lose in the quality of training or mentoring. Departmental programmes often produce more close-knit communities, with seminars, journal clubs or other events geared specifically to their graduate students.

Another structure is the bridge programme, which offers US students the chance to apply to a master’s programme that filters directly into a PhD programme on the same or a nearby campus. (The master’s-to-PhD route is common in the United Kingdom.) Such programmes are often a sound choice for those who feel that they need more preparation for doctoral studies. LaNell Williams, a second-year biophysics student, found that the Fisk–Vanderbilt Bridge Program run by Fisk and Vanderbilt universities

THE VALUE OF HINDSIGHT

What didn’t work for graduate-school applicants

PhD students who applied to programmes in the past several years share the wisdom and insights they’ve acquired from their own experiences, and offer words of warning.

Joseph Rodriguez, a postdoc at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, warns students to be aware of the specialized, standardized tests required for admission to certain programmes — and to prepare for them well in advance. He signed up at the last minute to take the physics Graduate Record Examination (required for most US physics PhD programmes), scored poorly and wound up in a programme in which no one was working on extrasolar planets, his research interest. After treading water for many months, he transferred to Vanderbilt University, all of which cost valuable time.

Priyanka Kothari, a third-year PhD student

in biochemistry, cellular and molecular biology at Johns Hopkins University in Baltimore, Maryland, counsels applicants to keep a close eye on their prospective department’s culture and make-up when they are on campus. During an interview visit at one university, she noticed graduate students introducing themselves to one another — a clear warning that the department had not fostered a sense of community and mutual support. “I want to do great science, but I also want a relationship with other faculty members and other students in the department,” says Kothari. “That networking is so critical for becoming the best scientist.” She was dismayed during another interview when she noticed that the programme’s faculty members were mostly white men. “Either the department doesn’t care enough to change

that or it doesn’t see it — both of which are red flags,” she says.

Allatah Mekile, a second-year PhD student at Johns Hopkins University in Baltimore, Maryland, says that location played a part in her decision as she weighed up whether she would feel comfortable living near the university. Considerations included how far she’d have to travel to get groceries, whether she’d feel safe leaving campus late at night and whether she could easily bring her car with her.

Rodriguez explains why students should think carefully when choosing programmes. “Graduate school is not just 5–6 years, but also working 50–60 hours per week and taking the hardest courses you’ve ever taken in your life,” he says. “You’re going to burn yourself out if you don’t really like what you are doing.” **K.P.**



The Fisk–Vanderbilt Bridge Program links up students and mentors.

FISK UNIV.

in Nashville, Tennessee, let her meet up with other students who were from groups that are under-represented in science. In contrast to her experience as the only woman of colour in her undergraduate physics studies, Williams says that after a year in the Fisk–Vanderbilt programme, she feels comfortable and has formed a community with fellow students. “I have been able to thrive,” she says, “and see myself as a physicist.”

The doctoral application process is not too early to think about ultimate career goals, says Horazdovsky. “Those can change,” he says. “However, you need to make sure you will have tools or experiences to achieve your goal by the end of graduate school.” For example, students who think they want to work at a mainly undergraduate institution will want significant teaching experience. Students who aim for industry will need exposure to business, companies and the jobs that PhD holders occupy. Students should also find out whether their programme of choice hosts, or at least encourages students to attend, conferences and workshops that help them to build teaching, networking and communications skills.

Many programmes include career-development components that give students real-world exposure to career tracks. These can be extremely helpful for students who are not aiming for an academic research position and can include university internships, external internships and other options.

TRUE GRIT

Students at the application stage need to stand out from the crowd to get accepted by their school of choice. David Charbonneau, director of graduate admissions for Harvard University’s department of astronomy in Cambridge, Massachusetts, looks for students who have persevered in the face of obstacles. “Most of what we do in science leads to dead ends,” he says. He seeks students who are passionate and hard-working, and who have demonstrated new ways of tackling

problems — for example, by working through solutions to an ambitious research problem for several years. These attributes should come across through concrete examples in their letters of recommendation, he says. McLean says applicants should personalize their application statements by including a paragraph explaining which faculty members within a programme they would like to work with, and why.

If prospective PhD students are unsure whether graduate school is the right decision, they should take a year or two to work as a research assistant in an academic or industry lab before making the hefty commitment to doctoral studies. Taking that time is no longer viewed as a negative, says McLean, but instead shows that applicants have realistic expectations and are aware of what’s ahead. Allatah Mekile was uncertain of her next steps after finishing college at East Stroudsburg University of Pennsylvania, so she moved home and took an entry-level position as a research associate at a supplier of nutritional products. There, she worked for two years on a metabolic-engineering project before applying to graduate programmes; she is now a second-year doctoral student in biochemistry, cellular and molecular biology at Johns Hopkins University in Baltimore, Maryland. She says that her experience in industry also helped her to explain in her application letter why and how certain programmes aligned with her career goals.

By eschewing the conventional path of going immediately into a doctoral programme after earning a bachelor’s degree, and gambling that she’d be better prepared, Mekile showed that she was ready for graduate studies, says Bogle. “The whole point of going to graduate school is to take a bit of a risk. If you want to play it safe all the way through, then maybe graduate school — or research — isn’t for you.” ■

Kendall Powell is a freelance science journalist based in Lafayette, Colorado.

GENDER BALANCE

Culture clash

Scientific disciplines that have a ‘masculine culture’ tend to deter women from pursuing those fields, a study finds (S. Cheryan *et al. Psychol. Bull.* <http://dx.doi.org/10.1037/bul0000052>; 2016). The study analysed 1,200 publications looking at women’s participation in science, technology, engineering and mathematics to learn why women are well represented in biology but not in physics, computer science and engineering. The authors found that the presence of negative stereotypes about women’s abilities and the lack of female role models were major factors in deterring women. But they also found that women who feared gender bias and discrimination might be more likely to avoid certain fields. Predictors of decreased participation included a lack of pre-university experience in the field and a lack of confidence. The low numbers are also linked more to a failure to recruit female students into the fields than in retaining them, suggest the authors. Creating a more inclusive culture is the best way to boost female participation, the authors say.

BIG PHARMA

UK drugs outsourced

Biopharmaceutical companies in the United Kingdom have cut research positions in drug discovery, according to a report released by the Association of the British Pharmaceutical Industry on 17 October, entitled *The Changing UK Drug Discovery Landscape*. In the past decade, almost all large UK drugmakers slashed in-house research jobs in discovery, the earliest stage of drug development, when researchers usually test hundreds of thousands of compounds to find one that could move into the next stage. Overall, there has been a net loss of several hundred positions. At the same time, however, large companies have increased their investment in drug discovery through outsourcing and collaborations. A number of UK contract research organizations (CROs) reported growth in partnerships with academic drug-discovery centres. Some CROs reported more drug-discovery employees, and about one-quarter of those reported staff increases of more than 25%. Yet some of the rise in CRO research jobs is also due to an increase in the number of contracts made with companies outside the United Kingdom, particularly in North America and the European Union, the report says. The UK biopharmaceutical industry employs more than 70,000 people.

BLOOD WILL TELL

Family ties.

BY TOM EASTON &
JACK MCDEVITT

Andy Pharon didn't know why he spent an hour every morning on FaceBook. Scandal! Outrage! Funny pussycats! More outrage! He might have been reading a tabloid, except that FaceBook was more respectable. Which mattered, as he was in Larry's. Martha came over. "Everything okay, Andy?"

"Excellent." He gave her his standard thumbs-up.

He was relieved moments later when his e-mail dinged. Sarah Mills, chief development officer at BioFutures Labs, wanted more ideas. Meeting at 10. Be there!

He finished his sweet roll and sipped his coffee. More ideas. He had nothing, but he couldn't say that, could he?

That was when the old guy with the roller bag squeezed between tables and stopped beside his chair. He was too well dressed to be a drifter but Andy still shook his head as he turned away for another sip of coffee.

"I thought I remembered this place," the guy said. "Came here every morning for five years."

Andy concentrated on his coffee cup and said nothing. Give 'em an inch, and they'll take a mile. Ten miles.

The guy looked down at him. "Hi, Andy. How's it going?"

"You know my name?"

"Sure. I'm you."

"What?" His face was lined and seamed, age spots, hardly any hair. Fifty years older than Andy. "Would you please go away?"

"We'll get time travel in about 30 years." He smiled. "I need a favour."

If this had been an e-mail, he would have hit delete. "Go away, gramps!"

The guy sighed. "I knew you would react that way. That I would. That I had. But I'm not a scammer. I don't want your money. And I already have your ID." He pulled out a chair and lowered himself into it. Then he produced a wallet. "See?"

Driver's licence. His picture with the name Andrew Pharon. Birth date was correct. Issue date: 2072. That would make him over 80.

Andy stared at him. The guy was smiling.

"What do you want?"

The smile faded.

"Some of your blood."

Andy sat frozen.

Had his life turned

into a vampire fantasy?

"Just some plasma, actually."

"Why?"

"Your people are already working on it. Putting young plasma into an old body can turn the clock back."

Andy nodded. If it was true ... "But why me?" Even as he spoke, he knew the answer. His own young plasma would work better than anyone else's. He really *was* a time-traveller.

Andrew grinned and delivered his standard thumbs-up, removing all doubt.

"Andy!" Martha waved at him. "You gonna be late!"

He waved back. This was one reason he liked Larry's. They cared.

The old guy was still sitting there, waiting for his response. But it was ridiculous. Time travel wasn't possible. "You have got to be pulling my leg."

The guy shook his head. "No. I just need a couple of pints today, and again next week and the week after." He looked at his bag. "The equipment's right here."

"I'm sure it is. But there's no way I'm letting you stick needles in me. And I've got to run." Andy tucked his tablet into his briefcase and stood.

"But ...!" He looked stricken, as if he had never dreamt that his own self would turn him down. "But I'm *you*! We're even closer than blood kin!"

"Pardon me. I have to leave." Incredibly, the guy was smiling as Andy went out the door.

Andy glanced over his shoulder and headed down the sidewalk, barely noticing the fumes of the remaining gasburners or the fragrance of the vagrant at the corner. The old guy wasn't following him. Thank God. Maybe he should switch coffee shops for a few days. But then the guy might just show up on his

doorstep. That would freak the hell out of his girlfriend.

Okay. Now he had to come up with an idea for Sarah.

BioFutures focused on the microbiome. Their last big success was a probiotic ointment for getting rid of acne. Lately they'd been working on figuring out how to manipulate bacteria in the gut to control obesity. They were close, which was why they needed new ideas. Had to keep the pipeline flowing.

Maybe the old guy had something?

Not time travel. But he recalled reading something about plasma and ageing. It wouldn't take long to check.

Once in the building, he went directly to his cube and started the search. And yes, they *were* working on it, testing it on people and making slow progress. The idea went back a century, when someone spliced the veins of a young mouse and an old mouse together. The old one got perkier, healthier, younger. The young one aged.

And plasma could be frozen.

He almost laughed.

It took him an hour to write the proposal. Start with some research into whether one's own young plasma is really better than a stranger's. Use mice, where the difference between young and old isn't great. If it checks out, then start collecting plasma, freeze it, store it, and when the donor turns into an old guy ...

He thought Sarah would like it. It was the perfect business plan, complete with references and links. Sell a promise, much like the old cryonics scam. Collect the money now, and worry later about whether the product actually works. Although this one seemed much more likely to be a success than cryonics ever had.

He would be among the very first to bank his plasma. And his older self knew how it had worked out. No wonder he'd sat there smiling when Andy walked out. ■

Tom Easton is a retired theoretical biologist who has written science-fiction novels and criticism and edited anthologies in addition to more academic work. Jack McDevitt is a prolific, award-winning novelist with an abiding interest in alien contact. A Philadelphia native, he has been, among other things, a naval officer, an English teacher and a management trainer for the US Customs Service.

➤ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm